UnIPaN: **Un**foldings and Abstract **I**nterpretation for **Pa**rametric Biological Regulatory **N**etworks

PhD proposal

Advisors

- Stefan Haar (DR Inria, Inria Saclay, Mexico team) http://www.lsv.ens-cachan.fr/~haar/
- Loïc Paulevé (CR CNRS, LRI, BioInfo team) http://loicpauleve.name

Context

The analysis of dynamics of Biological Regulatory Networks (BRNs), notably signalling and gene regulatory networks, faces the uncertainty of their computational model. Indeed, most of the knowledge available concerns the *existence* of (possibly indirect) interactions between biological entities (proteins, RNAs, genes), and rarely goes into the details on *how* different regulators of a same target cooperate, and even less on consistent rates for those interactions. In this regard, qualitative modelling approaches, such as Boolean and Thomas networks [3], offer an appropriate level of abstraction for these systems dynamics. As they are based on the interaction graph, they require few additional parameters, compared to classical quantitative models. Nevertheless, determining those *discrete* parameters is a well known challenge, and a major bottleneck for providing robust predictions from computational models.

The interaction graph of a BRN imposes dependencies in the evolution of each node: if only a and b regulate c, then the state of c is solely determined by them. But those dependencies do not suffice to specify the logical function of each node: c could require both a and b active in order to become active, or just one of them, or only a active and b inactive, etc. The parameters of a BRN are then the logical functions associated to each node, constrained by the interaction graph.

Parametric BRNs raise two challenges: (1) the identification of parameters for which the BRN satisfy given dynamical properties; (2) the analysis of the Parametric BRN's dynamics as such, for making predictions that are robust with respect to variability in the network. Parameter *identification* can be seen as a particular case of model synthesis [13], but in a very constrained setting, in the sense that the resulting model should be in accordance with the interaction graph.

The seminal work of Thomas and Thieffry [21], later formalized by Bernot et al [11], consists in enumerating all possible logical functions according to the interaction graph, and, for each instance, verifying its accordance with known dynamical properties. The major drawback of this approach is that the number of possible logical functions (parameters) is of high combinatorial complexity: in the case of Boolean networks, each node has 2^d parameters, leading to a search space of up to 2^{2^d} parameter values, where d is the in-degree (direct regulators) of the node.

Recent work attempts to provide more scalable approach than naive enumeration. or symbolic executions. In [9, 12], the authors rely on the exploration of the state space with symbolic encoding for tracking the parameters leading to each transition. Due to the explicit state space exploration, those methods are still limited in term of scalability. In [4], the authors use Hoare logic to deduce constraints on the parameters that can generate a given (partial) traces, with a limitation for infinite executions. In [16], the authors proposed a more scalable method, but which only over-approximates the set of parameters that allow to reproduce time series data.

So far, very few works have addressed dynamical analyses of a Parameteric BRN, i.e., where some of the parameters are unknown. In [18], the modelling framework allows to reason on the union of the dynamics of BRNs with different parameters. In such a framework, reachability properties can be computed on such

a union, but the allow only partial answers: if reachability does not hold in the union, then it cannot hold for any of the included BRNs; but if reachability holds, we cannot conclude. - With the same approach, [17] introduced an algorithm that can predict mutations for preventing a reachability to occur (cut sets) which would apply on any instance of BRNs compatible with the input Parametric BRN. But, again, those algorithms only give partial (albeit correct) predictions.

Objectives

The aim of this PhD thesis is to explore and develop the theory of concurrency and abstract interpretation for tackling the analysis of Parametric Boolean networks, and their discrete generalizations.

The scalability problems arising from concurrency can be harnessed by taking advantage of the parallelism of independent transitions allows touse reduce the dynamics to analyse by taking advantage of "parallel" transitions: if transitions t_1 and t_2 can be applied in a same state without interfering, instead of considering both sequences $t_1 \cdot t_2$ and $t_2 \cdot t_1$, one could provide a partial order in which t_1 and t_2 are not ordered. This has the effect of ignoring enumeration of interleavings. Arguably, the most powerful framework for expressing concurrency is the Petri net formalism [14], together with techniques for partial order unfoldings [15, 7, 8]. There, causal dependencies between transition occurrences are captured by structural relations. The representation of system dynamics via Petri net unfoldings is both more precise and, under exploitation of adequate cutoff techniques [8], considerablly more compact than explicite global state space models that necessitate to store and explore all sequential interleavings of concurrent actions. Exploring and using these structures is at the heart of the MEXICO team's activity (see e.g. [6, 20, 10, 1, 2]).

Now, BRNs are typically large (numerous nodes), but with sparse interactions: each node is regulated by a few other nodes compared to the size of the network. Therefore, BRNs can show a high degree of concurrency, which makes concurrency-aware techniques very promising for capturing their dynamics. The Mexico team at Inria/LSV and BioInfo team at LRI have initiated research in that direction, supporting the applicability and scalability of those approaches for actual biological networks [5].

Its combination with the abstract interpretation of automata networks, dynamics introduced by Paulevé et al [19] will be explored, aiming at providing compact, approximated, representations of Parametric BRNs dynamics, from which strict results on dynamical properties can be derived.

Workplan

The PhD thesis will develop two research axes:

- Semantics of the dynamics of Parametric Boolean networks (BNs) and generalizations for Petri nets. The goal is to produce representations of the possible traces of BNs with variability in their parameters. These representations will aim at sharing as much as possible (partially ordered) sequences of transitions that would be identical in BNs with different parameters. A first approach will focus on extending Petri net unfoldings where transitions are constrained by the domain of parameters. Then, further semantics of concurrency and abstractions of parameters and transitions will be explored.

- The design of algorithms for dynamical analysis that act on a Parametric BN. In particular, the project will focus on the characterization of reachable attractors in the parametric case, and of the cut sets for reachability. This research direction will first consist in defining the notion of attractor and of cut set in the parametric case, then study how they can be derived using unfolding and abstract interpretation. The objective is to provide correct and complete identification of attractors and cut sets and, conversely to be able to derive sets of parameters showing a given attractor/cut set.

Finally, the applicability of algorithms will be experimented on benchmark and actual models of biological regulatory networks, both from the literature and from joint project with partners in systems biology.

The theory developed in this thesis will have a strong impact on the current methods to address dynamical properties in biological networks: providing scalable algorithms to reason on models with variability in part of their specification is a key challenge for computational systems biology. At the same time, the investigation of concurrency will benefit from the extension of its field of investigation to systems biology, where the study

of discrete-event causal dependency and exploitation of independence is a recent and highly stimulating endeavor.

Scientific environment

The PhD student will be co-supervised by **Stefan Haar**, senior researcher INRIA and and head of the Mexico team in LSV, and by **Loïc Paulevé**, junior researcher (CNRS) in the BioInfo team in LRI.

MExICo is an INRIA project team located at the ENS Cachan's CNRS Laboratory LSV. The team specializes in formal methods that exploit the concurrent and interactive nature of complex systems, with an emphasis on dynamical system models such as variants of automata or Petri nets.

Stefan Haar works on exploration of causality in the behavior of concurrent systems, in particular using Petri net unfoldings, to solve supervision tasks such as conformance testing or fault diagnosis for partially observable systems. He directed the RNRT project SWAN (2003-2006) on autonomic telecommunication networks, and participated in the EU projects DISC (2007-2011) on Distributed Supervisory Control of Large Plants, HYCON 2 (2010-2014) on Highly-complex and networked control systems, and the ANR Project IMPRO (2010-2014) on implementability and robustness of real-time concurrent systems; he directed the DIGITEO/DIM doctoral project TECSTES on Testing Concurrent Systems using Event Structures.

The BioInfo team at LRI has a research activity focused on formal methods for capturing the transient dynamics of large-scale Boolean and discrete networks.

Loic Paulevé works on abstractions of causality in automata networks in order to design efficient algorithms for the formal analysis of reachability properties and for deducing potential mutations for controlling their emergence in biological networks. He was involved in the BioTempo ANR (2011-2014) project on computational methods for analysing biolo- gical network dynamics; and is involved in the HyClock ANR (2015-2017) project on the modelling and analysis of networks involved in the circadian clock and cell cycle for chronotherapies.

Cooperation. The Mexico team of LSV and BioInfo team of LRI have initiated a scientific collaboration in 2014 on the development and on the applications of Petri net unfoldings to models of biological networks [5]. Both teams are currently participating in the preparation of an ANR PRCI proposal between France and Luxembourg on algorithms for cell reprogramming, as well as in an INRIA associated team, LIFEFORM, with the University of Newcastle (UK) on formal methods for synthetic biology. S. Haar is now an associate member of the BioInfo team; L. Paulevé and he coordinate the DigiCosme working group TheoBioR on computational methods for biological networks.

Required background

The student is expected to have a strong background in logic, automata theory and algorithmic, as well as in computational systems biology, notably on Boolean network theory. Part of the work involves the implementation of a prototype for evaluating performance of the new algorithms, therefore advanced skills in programming (e.g., C++ or OCaml) are advised.

References

[1] Sandie Balaguer, Thomas Chatain, and Stefan Haar. Building occurrence nets from reveals relations. *Fundamenta Informaticae*, 123(3):245–272, May 2013.

[2] Paolo Baldan, Andrea Corradini, Barbara König, and Stefan Schwoon. McMillan's complete prefix for contextual nets. In *Transactions on Petri Nets and Other Models of Concurrency I*, volume 5100 of *Lecture Notes in Computer Science*, pages 199–220. Springer, November 2008.

[3] Gilles Bernot, Franck Cassez, Jean-Paul Comet, Franck Delaplace, Céline Müller, and Olivier Roux. Semantics of biological regulatory networks. *Electronic Notes in Theoretical Computer Science*, 180(3):3 – 14, 2007.

[4] Gilles Bernot, Jean-Paul Comet, Zohra Khalis, Adrien Richard, and Olivier Roux. A genetically modified hoare logic. *arXiv preprint arXiv:1506.05887*, 2015.

[5] Thomas Chatain, Stefan Haar, Loïg Jezequel, Loïc Paulevé, and Stefan Schwoon. Characterization of reachable attractors using Petri net unfoldings. In Pedro Mendes, Joseph Dada, and Kieran Smallbone, editors, *Computational Methods in Systems Biology*, volume 8859 of *Lecture Notes in Computer Science*, pages 129–142. Springer Berlin Heidelberg, 2014.

[6] Thomas Chatain, Stefan Haar, Maciej Koutny, and Stefan Schwoon. Non-atomic transition firing in contextual nets. In Raymond Devillers and Antti Valmari, editors, *Proceedings of the 36th International Conference on Applications and Theory of Petri Nets (ICATPN'15)*, volume 9115 of *Lecture Notes in Computer Science*, pages 117–136, Brussels, Belgium, June 2015. Springer.

[7] J. Esparza and K Heljanko. *Unfoldings - A Partial-Order Approach to Model Checking*. EATCS Monographs in Theoretical Computer Science. Springer, 2008.

[8] J. Esparza, S. Römer, and W. Vogler. An improvement of McMillan's unfolding algorithm. *Formal Methods in System Design*, 20:285–310, 2002.

[9] Emmanuelle Gallet, Matthieu Manceny, Pascale Le Gall, and Paolo Ballarini. An ltl model checking approach for biological parameter inference. In *Formal Methods and Software Engineering*, pages 155–170. Springer, 2014.

[10] Stefan Haar, Christian Kern, and Stefan Schwoon. Computing the reveals relation in occurrence nets. *Theoretical Computer Science*, 493:66–79, July 2013.

[11] Zohra Khalis, Jean-Paul Comet, Adrien Richard, and Gilles Bernot. The smbionet method for discovering models of gene regulatory networks. *Genes, Genomes and Genomics*, 3(1):15–22, 2009.

[12] H. Klarner, A. Streck, D. Šafránek, J. Kolčák, and H. Siebert. Parameter identification and model ranking of thomas networks. In *Computational Methods in Systems Biology*, pages 207–226. Springer Berlin Heidelberg, 2012.

[13] Maciej Koutny and Marta Pietkiewicz-Koutny. *Synthesis problem for petri nets with localities*. Springer, 2012.

[14] T. Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, April 1989.

[15] M. Nielsen and G. Plotkin, G.and Winskel. Petri nets, event structures, and domains, Part I. *Theoretical Computer Science*, 13:85–108, 1981.

[16] Max Ostrowski, Loïc Paulevé, Torsten Schaub, Anne Siegel, and Carito Guziolowski. Boolean Network Identification from Multiplex Time Series Data. In *CMSB 2015 - 13th conference on Computational Methods for Systems Biology*, volume 9308 of *Lecture Notes in Computer Science*, pages 170–181, Nantes, France, September 2015. Springer International Publishing.

[17] Loïc Paulevé, Geoffroy Andrieux, and Heinz Koeppl. Under-approximating cut sets for reachability in large scale automata networks. In Natasha Sharygina and Helmut Veith, editors, *Computer Aided Verification*, volume 8044 of *Lecture Notes in Computer Science*, pages 69–84. Springer Berlin Heidelberg, 2013.

[18] Loïc Paulevé, Morgan Magnin, and Olivier Roux. Refining dynamics of gene regulatory networks in a stochastic π-calculus framework. In *Transactions on Computational Systems Biology XIII*, volume 6575 of *Lecture Notes in Comp Sci*, pages 171–191. Springer, 2011.

[19] Loïc Paulevé, Morgan Magnin, and Olivier Roux. Static analysis of biological regulatory networks dynamics using abstract interpretation. *Mathematical Structures in Computer Science*, 22(04):651–685, 2012.

[20] Hernán Ponce de León, Stefan Haar, and Delphine Longuet. Model-based testing for concurrent systems with labeled event structures. *Software Testing, Verification and Reliability*, 24(7):558–590, November 2014.

[21] Denis Thieffry and René Thomas. Dynamical behaviour of biological regulatory networks—ii. immunity control in bacteriophage lambda. *Bulletin of Mathematical Biology*, 57:277–297, 1995.