

Title : Statistical Verification of Distributed Programs Within SimGrid

- **Location:** INRIA Nancy Grand Est research center --- LORIA, Nancy, France

- **Research theme:** [Algorithmics, Programming, Software and Architecture](#)

- **Project-team:** VeriDis

- **Scientific Context:**

The recent evolutions of High Performance Computing platforms (multicore and NUMA architectures, accelerators such as GPUs or Xeon Phi, complex interconnection networks, ...) have broken the traditional uniform parallel programming model. Hence, to exploit these new architectures, application developers have no other choice but to make their applications adaptive and to remove synchronizations as much as possible. To cope with resource heterogeneity and complex topologies, they also come up with sophisticated data management, scheduling, and load balancing strategies. This shift has two consequences. First, parallel programming has always been recognized as being much more complicated than sequential programming, and the need to resort to complex scheduling or communications strategies and to loosen synchronizations makes it even more error-prone. The resulting bugs occur almost never at small scale but quite frequently and in ways that are extremely hard to reproduce at large scale, which makes them particularly difficult to track down and remove. Second, the ever-growing complexity of software and hardware makes the understanding of their performance almost impossible. There are now dozens of layers with complex interactions that can be used to optimize performance. Correctly configuring such systems has become a nightmare even for experts.

SimGrid [1] is a software platform that provides models and APIs for simulating distributed systems; it has been shown capable of computing precise performance estimates for distributed programs running on various execution platforms. SimGrid also contains a model checking component [2] for dynamically verifying system correctness. This component aims at exhaustively verifying the possible system executions, even those that are unlikely to occur in practice, and thus at detecting so-called “Heisenbugs” that are extremely difficult to reproduce.

The main limitation of model checking as it is currently implemented in SimGrid is the so called “combinatorial explosion” problem that makes verification feasible only for small instances of systems. Moreover, knowing that some execution exists that violates a correctness property is sometimes not interesting. For example, in a large scale system, some messages transmissions can be delayed, and for such systems, even a 200ms delay (the typical TCP timeout) can have dramatic consequences in terms of performance as one processor being blocked can result in many others stopping as well, waiting for this processor’s data in order to proceed with their own computation. Instead of declaring that such an execution is possible, it would be more interesting if one could analyze the likeliness that it actually occurs. Doing so requires introducing probabilities in the system description, for example, on transmission delays. While probabilistic model checkers exist (e.g., [3]) and can in principle be used for analyzing such systems, they require a formal model of the system rather than C code as SimGrid, and they again suffer from the problem of combinatorial explosion.

- **Missions:**

The objective of this project is to adapt and implement the technique of statistical model checking [4] within SimGrid. In this approach, instead of attempting to exhaustively compute all possible system behaviors, executions are sampled according to the probability distributions associated with the target execution platform in order to estimate the probability that a property is satisfied in the system. Beyond the analysis of Boolean properties, the same technique can provide estimates for other relevant measures such as the average number of messages in a communication buffer or the expected time for achieving a task. Such a capability would be a useful complement to the current functionality of deterministic performance evaluation through simulation, by allowing for a range of behaviors instead of a precisely determined one. Because statistical model checking is based on sampling individual executions, there is no need to compute and store the entire state space of the system, and much more complex systems can be analyzed than is possible using traditional (probabilistic) model checking techniques (see e.g. [5]).

The concrete objectives of the proposed post-doctoral research are listed below. The range of subjects that will actually be covered will be determined taking into account the interests of the candidate and will be adapted according to the progress of the work. The subject combines conceptual research and implementation tasks to make statistical model checking available to users of the SimGrid platform, for applications to real-life programs and platforms.

Post-doctorants INRIA 2016

- Identify relevant probabilistic parameters and properties of interest supported by the existing SimGrid infrastructure.
- Implement a generic interface for performing statistical verification on top of SimGrid. A first experiment has been carried out within a master's thesis in 2015 and enables some simple analyses, using SimGrid as a black box. An interesting extension would be to follow the work done in COSMOS [5] to obtain and exploit information generated during the execution.
- Evaluate the scalability of the technique and the expressiveness of the extensions mentioned above by applying the approach to relevant examples that have already been implemented within SimGrid.
- Study a possible coupling with performance evaluation as traditionally performed within SimGrid.

- Bibliography

- [1] H. Casanova, A. Legrand, M. Quinson. *SimGrid: A Generic Framework for Large-Scale Distributed Experiments*. 10th EUROS/UKSim International Conference on Computer Modelling and Simulation (UKSim 2008), pp. 126-131. IEEE, 2008.
- [2] S. Merz, M. Quinson, C. Rosa: SimGrid MC: Verification Support for a Multi-API Simulation Platform. *Formal Techniques for Distributed Systems (FMOODS/FORTE 2011)*, LNCS 6722, pp. 274-288. Springer, 2011.
- [3] M. Kwiatkowska, G. Norman, D. Parker. PRISM 4.0: Verification of Probabilistic Real-time Systems. *Proc. 23rd Intl. Conf. Computer Aided Verification (CAV 2011)*, LNCS 6806, pp. 585-591. Springer, 2011.
- [4] A. Legay, B. Delahaye, S. Bensalem: Statistical Model Checking: An Overview. *1st Intl. Conf. Runtime Verification (RV 2010)*, LNCS 6418, pp. 122-135. Springer, 2011.
- [5] P. Ballarini, H. Djafri, M. Dufлот, S. Haddad, N. Pekergin. COSMOS: A Statistical Model Checker for the Hybrid Automata Stochastic Logic. *Eighth Intl. Conf. Quantitative Evaluation of Systems (QEST 2011)*, pp. 143-144. IEEE, 2011.

- Skills and profile

Required qualification: PhD in computer science

We expect the candidate to be familiar with the following subjects:

- Solid knowledge of formal verification techniques, in particular model checking and/or runtime verification.
- Solid programming experience, in particular with system programming in C.
- Basic knowledge of probability theory and statistics.

- Additional information:

Environment and contact:

The post-doctoral researcher will be hosted by the VeriDis research group of Inria Nancy – Grand Est and LORIA and will interact with the teams MEXICO of Inria Saclay and MYRIADS of Inria Rennes – Bretagne Atlantique and IRISA. The objective of the VeriDis team is to contribute research on the modeling and verification of distributed algorithms and systems. VeriDis was involved in the design and implementation of the existing model checker within SimGrid. We intend to intensify our involvement in making formal verification techniques to “real-world” programmers, and we believe that the combination of statistical model checking and simulation for performance evaluation will be fruitful. The statistical model checker COSMOS has been developed mainly by the team MEXICO, which has expertise in probabilistic systems and their verification. The MYRIADS team is specialized in designing and implementing systems and environments for autonomous service and resource management. It hosts core SimGrid contributors. All teams are partners in the Inria Project Lab HAC-SPECIS.

Contacts: Stephan Merz (stephan.merz@loria.fr) and Marie Dufлот-Kremer (marie.dufлот-kremer@loria.fr)

Additional links:

- <http://simgrid.gforge.inria.fr/>
- <http://veridis.loria.fr/>
- <http://www.inria.fr/en/teams/mexico>
- <http://www.irisa.fr/myriads/>

Duration: 18 months

Starting date: between Nov. 1st 2016 and Jan. 1st 2017

Salary: 2.620 euros gross monthly (about 2.115 euros net), medical insurance included

The required documents for an INRIA postdoc application are the following:

Post-doctorants INRIA 2016

- CV, including a description of your research activities (2 pages max) and a short description of what you consider to be your best contributions and why (1 page max and 3 contributions max); the contributions could be theoretical, implementation, or industry transfer. Include also a brief description of your scientific and career projects.
- The report(s) from your PhD external reviewer(s), if applicable.
- If you haven't defended yet, the list of expected members of your PhD committee (if known) and the expected date of defense (the defense, not the manuscript submission).
- Your best publications, up to 2.
- At least one recommendation letter from your PhD advisor, and possibly up to two other letters. The recommendation letter(s) should be sent directly by their author to the prospective postdoc advisor.

All these documents should be sent in two pdf files; one file should contain the publications, the other should contain all the other documents. These two files should be sent to your prospective postdoc advisor (in addition to the application on the web).

Additional benefits:

- Possibility of free French courses
- Help for finding housing
- Help for the resident card procedure and for husband/wife visa
- Lunch cost at Inria is 4,02 €