State Space Reduction for the Dynamic Formal Verification of Legacy Distributed Applications

Executive summary:
The state space explosion problem greatly hinders the applicability of exhaustive verification techniques. The goal of this work is to explore new reduction techniques specifically tailored to the dynamic verification of large legacy MPI applications.
Key skills required: Formal methods, model checking; Notions of system programming in C on Linux.
Research Unit: Inria Nancy – Grand Est, AlGorille team (leader: Martin Quinson)
Advisor: Martin Quinson — `martin.quinson@loria.fr` — `http://www.loria.fr/~quinson`
Paid internship, lasting 4 to 6 months in Nancy, France.

Context

SimGrid is a scientific instrument to study the behavior of large-scale distributed systems such as Grids, Clouds, HPC or P2P systems. It can be used to evaluate heuristics, prototype applications or even assess legacy MPI applications. Performance can be used through simulation while a dynamic formal verification tool called SimGridMC is included in the framework to evaluate the correction of protocols. The principle of this verification is to explore all the execution paths with an automated manner and check if each execution satisfied a given property. In contrast with model checking approach that require a model (either reconstructed automatically or manually built), the verification is performed on the real application through its controled execution. The application's model is thus unknown but implicitly explored.

As with model-checking, this verification process is hindered by the amount of execution paths to explore in any application. This well known problem is called the *state space explosion*. Even for small applications limited to a few processes, the number of explored states may reach several millions.

SimGridMC currently implements two reduction techniques. The first one called DPOR (Dynamic Partial Order Reduction) is based on the discovery of commutativity between independent actions. This independence is evaluate according to some theorems based on the communications semantic defined in the simulation framework. The second reduction is performed thanks to the detection of states already visited. The litterature presents several techniques in order to reduce this explosion but these are essentially practicable on abstract models of the studied applications, not on the real ones.

The goal of this internship is to study the applicability of existing reductions and for the verification of legacy applications, and then to provide new approaches inspired by the survey.

Proposed Work Plan

The work will be divided in several steps:

- A first work of bibliography may be necessary to extract a survey about others existing reduction approaches available both on abstract models and on real applications. A state of the art of some of them has been realized but their adaptability to our context remains unknown.
- Some of these methods should be implemented and compared with each other, thanks to our framework. A quality metric and the corresponding measurement tool should be proposed to evaluate the effectiveness of each reduction techniques, as well as their soundness and overall correction.

- This study should lead to the proposal of new reduction techniques specifically tailored to the context of MPI applications written in C/C++/Fortran. These new reductions will be tested on several applications, ranging from test cases to large applications counting hundreds of thousands of lines.

Skills required

The applicant should have a good understanding of formal methods in general. A previous exposure to the principle of model checking and the reduction techniques used in that context would be a plus.

To implement the proposed technique for evaluation, the applicant must be familiar with the C language under Linux, preferably with some knowledge of system or low level programming.

A previous exposure to the MPI protocol and its use for computationally intensive applications would be a plus, but is not mandatory at all.

References

- SimGrid: `simgrid.gforge.inria.fr`
- Talk on SimGrid, with a focus on verification: `www.loria.fr/~quinson/blog/2014/1016/CS2.pdf`
- Several publications on SimGridMC can be found on the web page of the tutor:
  - *SimGrid MC: Verification Support for a Multi-API Simulation Platform*, Cristian Rosa, Stephan Merz and Martin Quinson. 31st IFIP International Conference on Formal Techniques for Networked and Distributed Systems (FMOODS/FORTE), 2011.
  - *A Simple Model of Communication APIs – Application to Dynamic Partial-Order Reduction*, Cristian Rosa, Stephan Merz and Martin Quinson, 10th International Workshop on Automated Verification of Critical Systems (AVOCS), 2010.