Towards an Open Repository for Reproducible Performance Comparison of Parallel and Distributed Discrete-Event Simulators

Till Köster
University of Rostock
Rostock, Germany

Adelinde M. Uhrmacher
University of Rostock
Rostock, Germany

Philipp Andelfinger
University of Rostock
Rostock, Germany

ABSTRACT

Among the parallel and distributed simulation field's main subjects are the performance benefits of new methods and optimizations. However, performance evaluations of the various simulators often rely on custom models, parametrizations, and baseline implementations, which complicates direct comparisons. We present our vision and initial steps towards COMPADS, a benchmark model and repository for reproducibly comparing the performance of parallel and distributed simulators and their respective algorithms. COMPADS is short for COMparing Parallel And Distributed Simulators. The first results include a novel deterministic-by-design synthetic benchmark model inspired by PHOLD and La-pdes. The benchmark output is a checksum that attests to the correctness of an implementation and its execution. So far, implementations exist for the simulators ROOT-Sim and ROSS.

ACM Reference Format:
Till Köster, Adelinde M. Uhrmacher, and Philipp Andelfinger. 2022. Towards an Open Repository for Reproducible Performance Comparison of Parallel and Distributed Discrete-Event Simulators. In *SIGSIM Conference on Principles of Advanced Discrete Simulation (SIGSIM-PADS '22), June 8–10, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3518997.3534989

1 INTRODUCTION

One of the primary motivations for parallelizing discrete-event simulations is the reduction of execution times. The significance of novel methods and implementations developed with this motivation depends on their performance on well-defined benchmark problems relative to state-of-the-art implementations. Other computational fields follow this mode of evaluation. For instance, the various libraries that implement the Basic Linear Algebra Subprograms (BLAS) specification can be directly compared regarding their execution times for matrix multiplications. In contrast to the relative homogeneity of the dense vector and matrix operations defined by BLAS, the models executed using parallel and distributed simulation (PADS) can differ immensely in their computation and communication behavior, complicating fair comparisons across simulators.

ACM SIGSIM PADS '22, June 8–June 10, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9261-7/22/06.
https://doi.org/10.1145/3518997.3534989

A degree of comparability of results has been achieved by the use of synthetic benchmark models such as PHOLD [3], EPHOLD [1] and La-pdes [2] or proposals targeted towards specific simulation approaches such as DEVStone [4]. Further, as an essential step towards identifying parametrizations that capture the attributes of real-world problems, some recent work aims to characterize the performance-affecting attributes of a wide range of discrete-event simulation models.

Still, the various model variants and parametrizations used in performance evaluations of simulators make comparisons across studies difficult. Further, commonly employed benchmark models permit comparisons of the simulation results across simulators only on a statistical level.

2 BENCHMARK MODEL

Aiming to facilitate insights into the relative performance of the various simulators and to accelerate the propagation of "best-in-breed" methods, we describe our vision and initial results towards COMPADS (parametrizable and deterministic/parallel and distributed simulation benchmark), a benchmark model and open repository for reproducible performance evaluation of parallel and distributed simulators. We postulate that for the project to be beneficial to the community, two main requirements must be satisfied:

Reproducibility: We aim to enable researchers to obtain comparable results when repeating published experiments, even across programming languages, hardware platforms, and simulators. To this end, we propose a fully deterministic synthetic benchmark model related to PHOLD and La-pdes. We achieve this strict level of reproducibility by handling pseudo-randomness and simultaneous events as part of the model, which prevents deviating results among simulators. At termination, a single checksum attests to the correctness of the execution.

Mathematically speaking, a simulation based on our benchmark model can be viewed as a hash function on the model configuration. This hash function comprises numerous interconnected deterministic random number generators. Each entity is in itself a random number generator. When receiving an event, the local state is altered by the event payload, such that future random numbers also change. In effect, the overall system is comprised of communicating deterministic random number generators that alter one another's state.

We consider the model deterministic if it produces the same results 1. across multiple runs using the same simulator, and 2. across simulators, which poses two main challenges: Firstly, the generated pseudo-random numbers must always be produced using the same algorithm, and the assignment of random number streams to entities must be identical. Secondly, simultaneous events, i.e., multiple events that happen to occur at the exact same point in logical time, must be handled in a deterministic fashion.

Even with a high-quality random number generator and a time representation using double-precision floating point numbers, timestamp collisions occur with non-zero probability. Several existing works consider the ordering among simultaneous events using a tie-breaking policy [5]. However, simulators may vary in their tie-breaking policies, leading to divergent results on simultaneous events. For this reason, our benchmark model includes a mechanism that circumvents the need for tie-breaking, guaranteeing deterministic behavior even in the presence of simultaneous events.

Representativity: Performance evaluations should capture the performance properties of a reasonably comprehensive set of models relevant to practitioners without putting excessive weight on particular properties. To achieve this, the proposed benchmark model is configurable for various essential performance-affecting characteristics such as workload uniformity, computational granularity, message size, among others. To determine the parametrizations to be included in automated simulators comparisons by default, our project going forward depends on community feedback.

REFERENCES

[1] VAM Bonnet. 2017. *Benchmarking parallel discrete event simulations.* Master's thesis. Utrecht University.

[2] EunJung Park, S. Eidenbenz, N. Santhi, G. Chapuis, and B. Settlemyer. 2015. Parameterized benchmarking of parallel discrete event simulation systems: Communication, computation, and memory. In *2015 Winter Simulation Conference*. 2836–2847. https://doi.org/10.1109/WSC.2015.7408388

[3] Richard M Fujimoto. 1987. *Performance Measurements of Distributed Simulation Strategies.* Technical Report. Utah University.

[4] Ezequiel Glinsky and Gabriel Wainer. 2005. DEVStone: a benchmarking technique for studying performance of DEVS modeling and simulation environments. In *Ninth IEEE International Symposium on Distributed Simulation and Real-Time Applications*. IEEE, 265–272.

[5] Robert Rönngren and Michael Liljenstam. 1999. On event ordering in parallel discrete event simulation. In *Proceedings of the 13th Workshop on Parallel and Distributed Simulation.* IEEE, 38–45.