Scalable Trace Driven Parallel Network Simulation

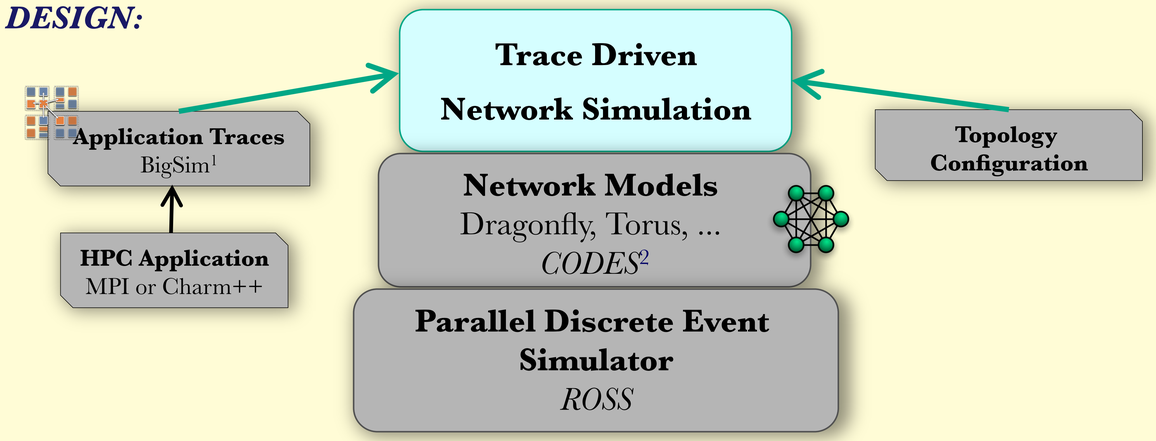
Bilge Acun[1,2], Nikhil Jain[1,2], Abhinav Bhatele[2], Laxmikant V. Kale[1]
[1]University of Illinois at Urbana-Champaign, [2]Lawrence Livermore National Laboratory

ABSTRACT: Simulation of HPC applications has an important role in performance prediction, studying different interconnect topologies and development of new architectures. The goal of this project is to integrate traces of production MPI and Charm++ applications with the CODES[2] network simulator in order to simulate production applications in addition to synthetic communication patterns. The CODES simulator provides network models for various topologies (such as torus and dragonfly) and is built on the scalable ROSS[3] Parallel Discrete Event Simulator (PDES) framework.

MOTIVATION:

- **State of the art:** Current simulators are not scalable or do not simulate production applications
  - Not scalable because they are not running in parallel, e.g. BigSim
  - Only simulate synthetic hard-coded communication patterns, e.g. CODES
- **Goal:** Creating a scalable, parallel, packet-level network simulator for production HPC applications
  - Get the task dependencies from the applications via traces and simulate their behavior

DESIGN:



Application Traces:

- In BigSim trace file format
- Records every event throughout the application
- Generated via BigSim Emulator
- Capable of emulating multiple MPI ranks per process.

Sample Trace for one processor
Time Stamp, Task ID, Name, Duration…Back&Forward Dependencies

```
-1.000000 47 AMPI_Bcast--time:5960 0.000006  ...  $B 46  $F 53
0.001148 48 start-broadcast--time:0 0.000000 ... $B  $F 49
-1.000000 49 AMPI_generic--time:3099 0.000003 .. $B 48 $F 50 52
-1.000000 50 end-broadcast--time:0 0.000000  ... $B 49  $F
0.001151 51 msgep--time:953 0.000001   … $B  $F
0.001154 52 RECV_RESUME--time:953 0.000001 ... $B 49  $F 53
-1.000000 60 user_code--time:0 0.000000 ... $B 59 54  $F 61
```

Network Models:

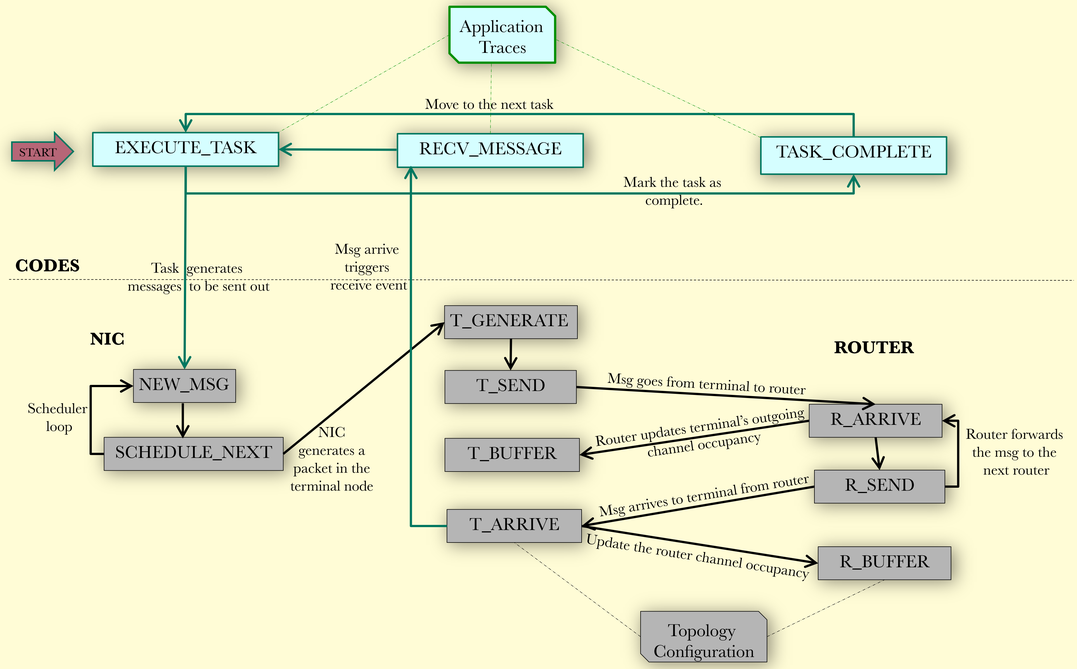
- Detailed models that CODES simulator provides
- Able to simulate 50 million nodes
- Built on ROSS Parallel Discrete Event Simulator

Dragonfly: IBM PERCS Model with Adaptive and Static Routing
Torus: IBM BG/P, IBM BG/Q
Other Simple Network Topologies

Parallel Discrete Event Simulator:

- ROSS: A scalable framework for PDES
  - Uses the Time Warp synchronization protocol
  - Can process billions of events per second
  - Has sequential, conservative and optimistic modes

EVENT FLOW:



SUMMARY and FUTURE WORK:

- Integrating production application simulation with CODES simulator
- Future work:
  - Add support for optimistic mode, current support is for conservative mode
  - Do a performance study of different simulators, CODES, BigSim, SST etc.

References:
1. BigSim, Simulating PetaFLOPS Supercomputers http://charm.cs.uiuc.edu/research/bigsim
2. CODES, Enabling CO-Design of Exascale Storage Systems, Argonne National Laboratory http://www.mcs.anl.gov/projects/codes/
3. ROSS, Rensselaer's Optimistic Simulation System, Rensselaer Polytechnic Institute https://github.com/carothersc/ROSS/wiki

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This work was funded by the Laboratory Directed Research and Development Program at LLNL under project tracking code 13-ERD-055 (LLNL-POST-658145).