UNIVERSITY of HOUSTON

PARALLEL SOFTWARE TECHNOLOGIES LABORATORY PSTL

Tuning of OpenMPI parameters

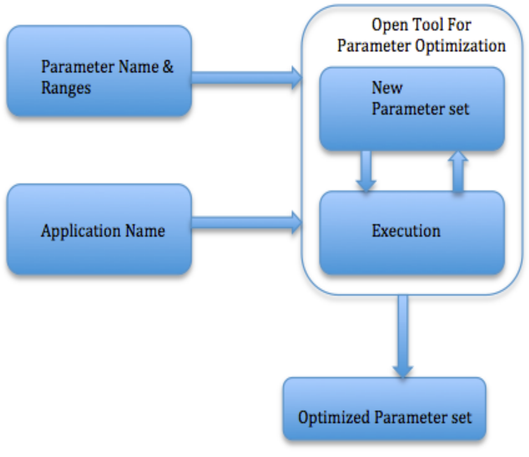
Shweta Jha and Edgar Gabriel
Parallel Software Technologies Laboratory, Department of Computer Science,
University of Houston, Houston, TX

Introduction

- Communication costs have a high impact on scalability of parallel applications
- Internal parameters of MPI library influence communication costs
- No default settings of parameters will be optimal for all the use cases and platforms
- Solution: Parameters of communication library need to be adjusted on a case by case basis

Open Tool for Parameter Optimization (OTPO)

- Open MPI specific tool to tune MCA parameters
- OTPO input:
  - ✓ List of parameters to be tuned
  - ✓ Parameter values to be explored
  - ✓ Benchmark name
- OTPO output
  - ✓ Parameter sets leading to minimal execution time of the benchmark



OTPO search strategies:

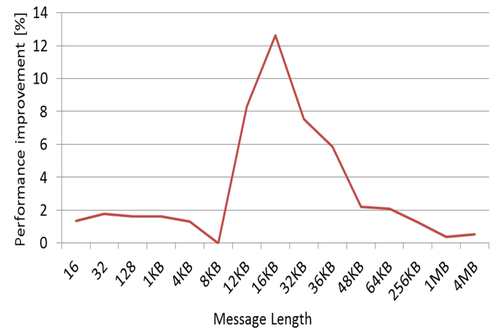
- Brute force search strategy
  - ✓ Tests all possible combination
  - ✓ Guarantees finding best possible combination
  - ✓ Long selection phase
- Attribute based search strategy
  - ✓ One attribute is optimized, while others are constant
  - ✓ Based on assumption that attributes are independent
  - ✓ Order of attributes is important
- 2K Factorial search strategy
  - ✓ Reduces number of combinations evaluated
  - ✓ Provides weight of each parameter on the performance
  - ✓ Quick output

Semi-automated parameter tuning

1. Identification of MPI functions and most frequently used message lengths in an application using a simple profiling library
2. Tuning of influential MCA parameters based on the results of step no. 1
3. Optimal parameter set is stored in PostgreSQL database
   - ✓ Parameter values + metadata
     - Host key: Identifier of the cluster
     - Application key: Key for a specific application
     - Application characteristic: characteristics pertaining to a particular case
4. Retrieval of optimized parameter set through mpiexec

Use case 1: Sensitivity of network parameter to message length

NetPIPE benchmark showed the maximum sensitivity for 12kB to 48kB message length.



Use case 2: Impact of optimized openib parameters on collective operations

Linear and Pairwise algorithm showed the performance improvement in the range of 20% - 60%.

Algorithm	no. of procs.	Default Exec. Time [ms]	Tuned Exec. Time [ms]	Relative Gain [%]
All-to-all linear	32	40.7	27.1	33.2%
All-to-all pairwise	32	31.9	25.1	20.6%
All-to-all bruck	32	56.6	56.3	0.4%
All-to-all linear	128	918.6	320.9	64.4%
All-to-all pairwise	128	35.9	26.5	25.9%
All-to-all bruck	128	57.8	57.7	0.2%

References:

Shweta Jha, Edgar Gabriel, Saber Feki,'Personalized MPI Library for Exascale application and environments', Workshop on Exascale MPI 2014 @ SC14, At New orlans, LA, USA