Impact and Limitations of Point-to-point Performance on Collective Algorithms

Shweta Jha and Edgar Gabriel
Department of Computer Science, University of Houston, Houston, TX

Improving the scalability of parallel applications by tuning the communication among the processes

Introduction

• Communication cost have high impact on scalability of parallel applications

• Internal parameters of MPI (also called Modular Component Architecture/MCA parameters) library influences the communication cost

• We aim to improve the theoretical and practical understanding on potential improvements of an application through the tuning process

Open Tool for Parameter Optimization (OTPO)

• OpenMPI specific tool to tune MCA parameters

OTPO Inputs:

✓ List of parameters to be tuned
✓ Parameter values to be explored
✓ Benchmark or application name

• Output:

✓ Parameter sets leading to least execution time

Fig 1: OTPO workflow

Prediction of improvement in collective operations

• Calculate the improvement in point to point operation for each message length using OTPO

• Calculate the expected improvement of collective operations for each message length using the formulas.

• Compare the performance

Equations for predicting the improvements:

• Allgather Recursive Doubling Algorithm:

Performance Models

Predict the expected improvement of a collective operation based on the improvement observed for a point-to-point communication using Hockney's model and the LogGP model.

The time to send a message of size m between two processes are given by

• Hockney's model:

\[ t = l + m/b \]

\( l \) - network latency; \( b \) - network bandwidth

Parameter to be improved: \( l \) and \( b \).

• LogGP model:

\[ t = L + 2o + (m - 1)G \]

\( L \) - delay; \( o \) - overhead; \( G \) - gap per byte

Parameter affected by tuning are \( G \) and \( o \).

Evaluation

• Broadcast Binary Algorithm:

Fig 2: Performance improvement in point to point communication operations (InfiniBand)

Fig 3: Performance improvement (actual and predicted Hockney’s and logGP) of Broadcast binary algorithm

Fig 4: Performance improvement (actual and predicted Hockney’s and logGP) of Allgather Recursive Doubling algorithm

Conclusions

• Many algorithm predict lower improvement than point to point improvement

• Predictions matches the actual measurements for a number of algorithms

• Identification of two reasons why predictions and actual measurements might differ, namely

  a) sensitivity of some parameters to the number of processes used regarding their performance implications, and

  b) the fact that some collective operations use internally different message sizes than what has been tuned for

References