Assessment

Purushotham V. Bangalore
Department of Computer Science
University of Alabama

Assessment Goals

- Identify applications and input decks that are representative of ongoing and future communication challenges,
- Understand the qualitative and quantitative needs of these applications,
- Assemble public test suites based on this assessment to drive future center and broader community research on optimizing HPC communication systems

Application Characteristics

- Capture communication challenges faced by the broad range of DOE applications on Exascale architectures
- Identify MPI primitives and their common usage, including point-to-point, collective, and one-sided communication
- Represent inter-node communication patterns used by realistic applications, including neighbor halo exchange, static irregular, and dynamic irregular communication
- Express code complexity of full application frameworks and simplified proxy and mini-applications
- Represent relevant programming frameworks and languages
- Cover NNSA-relevant application areas

Prior Work

- Analysis of ECP proxy application suite (both static and runtime analysis)
  Sultana, N, Rüfenacht, M, Skjellum, A, Bangalore, P, Laguna, I, Mohror, K. Understanding the use of message passing interface in exascale proxy applications. *Concurrency & Computation: Practice & Experience*. 2021; 33:e5901. https://doi.org/10.1002/cpe.5901
- 100+ open source MPI applications (static analysis)
  Ignacio Laguna, Ryan Marshall, Kathryn Mohror, Martin Ruefenacht, Anthony Skjellum, and Nawrin Sultana. A large-scale study of MPI usage in open-source HPC applications. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (*SC '19*). Association for Computing Machinery, New York, NY, USA, Article 31, 1–14. https://doi.org/10.1145/3295500.3356176

ECP Proxy Apps Analysis: Findings

- Applications use a small subset of MPI, with significant overlap of those subsets
- Approximately half of the proxy apps use a hybrid (MPI + OpenMP) programming model. However, none of them use the MPI_THREAD_MULTIPLE mode
- MPI point-to-point calls dominate the communication in terms of the call count while MPI collectives occupy a significantly larger portion of the communication time as compared to point-to-point
- MPI_Allreduce is the most frequently used collective call (applications mostly use short messages)

Open Source HPC Apps Analysis: Findings

- Majority of the applications use a small subset of the MPI standard
- Large percentage of the applications do not use any of the advanced features in MPI (e.g., 67% of these applications use blocking send/recv calls)
- Most applications use MPI-1.0 features
- 2/3 of the applications use hybrid programming (OpenMP being the most popular choice)
- C++ is the dominant language used by these applications

Application Selection Process

- Narrowed this down to 40 DOE production applications and proxy or mini-applications
- Further discussion with TST team and NNSA lab collaborators to understand lab application needs
  - Identify production applications with communication challenges
  - Identify proxy or mini-applications that are representative of these production applications

Production & Proxy/mini Apps, I

Application Miniapp	Lab	Availability	Priority	Description
LLNL CFD/Mechanics Apps	LLNL	EC	2	ALE3D, Nike3d, etc.
Comb		Open	1	GPU stencil communication mini-application
EMPIRE	SNL	EC	3	Hybrid PIC Electrodynamics framework
EMPIRELite	SNL	In devel	2	Forthcoming EAR99 version
RefMaxwell	SNL	Open	1	Trilinos AMG solver used in EMPIRE
ExaMPM	ECP-COPA	Open	1	Proxy for EMPIRE particle push
HIGRAD	LANL	EC	1	Shock CFD application
Fiesta	UNM	Open	1	Partial open reimplementation & modernization of key HIGRAD features
MERCURY	LLNL	EC	3	Monte Carlo Radiation Dynamics
Quicksilver	LLNL	Open	2	Simplified Monte Carlo Transport Proxy

Production & Proxy/mini Apps, II

Application Miniapp	Lab	Availability	Priority	Description
PARTISN	LANL	EC	3	Neutron transport; DesignForward traces available
SNAP		Open	2	Proxy of PARTISN compute/communication patterns
SPARC	SNL	EC	3	Reacting and non-reacting hypersonic CFD code
MiniAero		Open	2	Unstructured Navier-Stokes solver mini-app
xRage	LANL	EC	3	Radiation Transport/Hydrodynamics Framework
CLAMR		Open	1	2D Cell-based adaptive mesh refinement mini-app.
EAP Proxy		In devel	2	Forthcoming xRage proxy application

Assessment Plan

Each phase involves modeling, evaluating, and optimizing applications that:

I. use predominantly point-to-point and nearest-neighbor communication on both CPU and GPU-based systems
   a. regular halo communication – Comb, Fiesta
   b. irregular halo communication – CLAMR, HYPRE

II. use predominantly collective communication primitives

Regular Halo Assessment Findings

- Large-scale GPU applications can spend 50% or more of their time in MPI communication overheads
- Optimization of communication activities (e.g., careful ordering and pipelining of packing loops, CPU-GPU copies, CPU-GPU synchronization, and ordering of MPI sends and receives) can reduce communication overheads by a factor of three and improve application performance by 30% or more
- The currently used MPI communication primitives place the burden of fine-grain communication system optimizations on the application programmer and higher-level communication abstractions that address this issue are required

See Poster: MPI Communication Performance of a Shock Hydrodynamics Application, Ryan Goodner, UNM

Irregular Halo Assessment Findings

- The performance varies drastically across MPI implementations (e.g., On Lassen Spectrum MPI greatly outperforming MVAPICH2-GDR)
- MPI communication performance on irregular communication patterns, particularly in the face of load imbalance varied significantly between MPI implementations, and can reduce the performance of some codes by 25% or more

See Poster: Communication Requirements and Challenges in Irregular Applications, Tanner Broaddus, UTC

Next Steps

- Assessment of application performance using higher level abstractions and new APIs
- Assessment of predominantly collective communication applications
- Assessment of more complex communication patterns in production NNSA applications

Questions