Approach Developed for Automatically Characterizing Parallel Application Communication Patterns

**Achievement:** Developed an approach for automatically recognizing and concisely representing the communication patterns of parallel applications that use the Message Passing Interface (MPI).

**Significance and Impact:** Characterizing parallel application communication patterns requires considerable expertise but greatly simplifies tasks such as proxy application validation, performance problem diagnosis, and debugging. Our automated approach significantly reduces the expertise necessary for effective characterization.

**Research Details:**
- Developed an approach that uses automated search and a communication pattern library to automatically recognize unknown communication patterns in MPI-based applications.
- Developed a concise representation of the parallel application communication patterns recognized by our characterization approach.

**Sponsor/Facility:** Funding for this work was provided by the Office of Advanced Scientific Computing Research, U.S. Department of Energy. The work was performed at Oak Ridge National Laboratory (ORNL).

**PI and Affiliation:** Philip C. Roth, ORNL


**Overview:** To enable characterization of parallel application communication patterns by non-experts, we developed an approach for automatically recognizing and parameterizing communication patterns in MPI-based applications. Beginning with a communication matrix that indicates how much data each process transferred to every other process during the application’s run, we use an automated search to recognize communication patterns within this matrix. At each search step, we recognize patterns from a pattern library in the communication matrix. Using a technique similar to astronomy’s “sky subtraction,” when we recognize a pattern we remove it from the matrix and apply our recognition approach recursively to the resulting matrix. Because more than one pattern might be recognized at each search step, the search produces a search results tree whose paths between root and leaves represent collections of patterns recognized in the original matrix. The path that accounts for the most of the original communication matrix’s traffic corresponds to the collection of patterns that best explains the application’s communication behavior. We implemented our approach in a tool called AChax that was highly effective in recognizing the communication patterns in a synthetic communication matrix and the regular communication patterns in matrices obtained from the LAMMPS molecular dynamics and LULESH shock hydrodynamics applications.

![Figure 1: Recognizing and removing the contribution of a 2D nearest neighbor pattern in a synthetic communication matrix. This represents one step in our automated search-based approach.](image1)

![Figure 2: Results tree produced by search-based automated pattern characterization. Nodes are labeled with the amount of communication yet to be explained, and edges are labeled with the recognized pattern. The path in red indicates the patterns that best explain the original communication behavior.](image2)

![Figure 3: Concise representation of recognized communication patterns for a LAMMPS run with 96 processes.](image3)