

Project Title: Overlay Transit Networking for Scalable, High Performance Data Communication Across Heterogeneous Infrastructure

Principle Investigator: Micah Beck, University of Tennessee

Main DOE Collaborators: Anthony Mezzacappa, Scott Klasky, Jonghyun Lee

Project's Goals

The project's primary goal is to integrate Logistical Networking with the tools and services that support wide area data intensive science application communities. A secondary goal is to enable the use of computation at network intermediate nodes to support application data management.

Description of Major Research Activities

1. Porting of Application I/O Libraries to Logistical Networking

Based on Unidata's NetCDF, LoCI Lab has released NetCDF/L which adds Logistical Networking capabilities. Like the traditional NetCDF, NetCDF/L can store data on a local filesystem and it can also store data on the global logistical network. By specifying a local filename or a Logistical Runtime System (LoRS) URL (i.e. lrs://), the user controls where the data is stored. NetCDF/L (3.6.0 and higher) supports CDF version 2 (CDF2) which supports files up to 8 exabytes (EB). A similar port of NCSA's HDF5 library will be completed soon, and will provide compatibility for applications written to that interface.

The libxio library, developed in collaboration with the DiDaS project in the Czech Republic, provides a standard Unix I/O interface (i.e. open(), close(), read(), write(), etc.) to access local files as well as logistical technology-based "network files" (exNodes). Researchers can port their Unix I/O applications to add logistical networking capabilities by adding 12 lines of code and recompiling. Like LoRS, libxio supports 64-bit file offsets, user configurable multi-threading, and more.

2. Libraries Available Through ROMIO Support of LN

Jonghyun Lee of Argonne National Laboratory, in collaboration with LoCI Lab, has adapted the underlying storage or "abstract device" layer of the MPI-IO parallel I/O library to use Logistical Networking infrastructure and middleware directly. This allows application programs to read and write data to local or remote network storage resources without modifying the interfaces or programming techniques that they currently use. The vehicle for this work is the ROMIO implementation of MPI-IO developed at Argonne, and it forms the foundation of a further stack of application middleware for parallel I/O, including the Parallel HDF5 and Parallel netCDF libraries. Logistical Networking implementations of these widely-used interfaces are direct consequences of

the underlying port of MPI-IO, as are any other systems or libraries based on ROMIO's "abstract device" layer or on the MPI-IO interface. This type of integration of innovative wide area storage technology with parallel I/O programming techniques developed for large data centers represents an important synthesis that will ensure that advances in infrastructure have the desired impact on the computational science community.

3. Indirect Management of Data through Storage Metadata

The NetCDF/L, HDF5/L, libxio, and POSIXIO/L libraries all support I/O to and from a single computer or a single node in a cluster or supercomputer. When researchers run processes across multiple computers/nodes, the result may then be a dataset spread distributed across a configuration-specific set of files. An application may then access the dataset in some other configuration, either as a single file or distributed differently to correspond to the topology of a different computing system. We are investigating providing a library to support the management and remapping of data through the manipulation of exNodes metadata, without requiring any movement of the data itself.

4. Performing Operations on Distributed Data Sets Stored at Depots

We have developed a mechanism for installing and allowing execution of limited operations on IBP depots, as well as two sandboxed execution environments (native machine code & JVM) to support mobile code. Initial applications of this capability include support for scientific visualization and data reorganization and rewriting.

5. Caching for libxio, POSIXIO/L, NetCDF/L and HDF5/L

We are investigating the use of automatic pre-fetching and caching of remote data to improve performance of applications that use these libraries. The appropriateness of various caching schemes must be validated with application behavior. Initial applications will be in data analysis and visualization.

7. Tape Archive Support

At Vanderbilt's Advanced Computing Center for Research and Education, led by Paul Sheldon, and in collaboration with LoCI Lab, a Logistical Storage (L-Store) element is being developed that uses an IBP depot to manage storage across a disk and tape archive system. The L-Store is intended as a storage element for the Open Science Grid, and it implements the Storage Resource Manager (SRM) interface which is a standard within the community. The L-Store will be used by ACCRE as the basis for a Bandwidth Challenge entry at SC05.