

## Distributed Technologies for Nuclear Fusion Data

S. G. Shasharina<sup>1</sup>, R. Eger<sup>1</sup>, J. R. Cary<sup>1</sup>, T. W. Fredian<sup>2</sup>, D. Alexander<sup>1</sup>

<sup>1</sup>Tech-X Corporation, 5541 Central Ave, Suite 135, Boulder, CO 80301, USA

<sup>2</sup>MIT Plasma Science and Fusion Center, 167 Albany Street · Cambridge, MA 02139 USA

### Abstract

Simulations in the nuclear fusion community are generating large data sets at remote sites. Visualization and analysis of these data sets is then difficult, as downloading it to local computers can take hours over even the fastest network connections. Problems such as these are what Grid technologies,<sup>1</sup> in particular the Globus Toolkit, are addressing. Until now nuclear fusion has been using non-Grid and domain-specific solutions for distributed data storage, acquisition and analysis. In the described work, we will compare performance, flexibility and ease of use of Globus and MDSplus,<sup>2</sup> as data transfer mechanisms, and HDF5<sup>3</sup> and MDSplus, as data formats. As a first step, we have done performance tests of data transfer in two models: one using the Globus Toolkit (GridFTP transfer of HDF5 files), another using MDSplus (mdsip transfer of equivalent MDSplus SIGNALS). To make these tests possible, an HDF5-to-MDSplus importer was developed. The preliminary tests on a 100Base-T Local Area Network show that one-stream GridFTP is 2-2.5 times faster than MDSplus, when large data sets are being transferred. When the size of the data is relatively small (<4.5 MB) MDSplus is more efficient. Tests with partial extraction and transferring of data indicated that MDSplus is more efficient in data extraction than the other scenario, which had a significant overhead associated with the Globus Resource Allocation Manager (GRAM), which we used for HDF5 dataset extraction. Similar to our LAN results, preliminary Wide Area Network transfer tests show that the GridFTP becomes more efficient for larger files (>45MB). We will present the tests results and discuss our recommendations for implementation of fusion data grids, which will hopefully combine the benefits of all the technologies: MDSplus, Grids and HDF5.

The work was supported by a DOE/SBIR grant DE-FG03-02ER83555.

<sup>1</sup> I. Foster, C. Kesselman, S. Tuecke, *The Anatomy of the Grid: Enabling Scalable Virtual Organizations*, International J. Supercomputer Applications, **15**(3), 2001.

<sup>2</sup> T. Fredian and J. Stillerman, *MDSplus Remote Collaboration Support – Internet and the World Wide Web*, Fusion Engineering and Design **43**, (1999).

<sup>3</sup> <http://hdf.ncsa.uiuc.edu/HDF5/>.