Large-scale FMS Simulations and Transcontinental Data Management and Analysis

Dan M. Davis and Ke-Thia Yao
ddavis@isi.edu and kyao@isi.edu

Information Sciences Institute
University of Southern California
Marina del Rey, California

The transformation of the U.S. Armed Forces to meet the challenges of the 21st Century presents new obstacles to the analysts and new opportunities for the Computational Scientists. The emphasis on urban, low-collateral damage, surgical, and decisive warfare has presented the U.S. DoD with the conundra of

- how to train, mission plan, and practice assaults in densely-populated civilian areas,
- how to assess new technology before it is produced, and
- how to effectively meet challenges that are not as yet made evident.

All of this needs to be done, all the while maintaining a low profile to keep from disturbing the peace of the various environments and to avoid panic among the various peoples that will likely constitute the future combat environment. FMS is an area, exclusive to the military, where computational science can assist by implementing large-scale, realistic, easily controlled and valid urban simulations, populated by hundreds of thousands of independent agent friendly forces, enemy forces, irregular forces, and civilians. These simulations can be richly enhanced with phenomenology such as weather, blast effects, daylight, chemical plumes, and other salient features of the battlespace of the next few decades. One such project is the Joint Experimentation on Scalable Parallel Processor (JESPP). There is a well defined and critical need to produce quantifiable results from simulations to support transformational findings for the Joint Forces Command (USJFCOM). This is driving the creation of very large and geographically dispersed data collections. The Joint Experimentation Directorate (J9) of USJFCOM and the Joint Advanced Warfighting Project is conducting a series of Urban Resolve experiments to investigate concepts for applying future technologies to joint urban warfare.

The recently concluded phase I of the experiment required, utilized and integrated multiple scalable parallel processors (SPP) sites distributed across the United States, enabled and hosted by the supercomputing centers at Maui (NHPCC) and at Wright-Patterson (ASC-MSRC) on a net including J9 at Suffolk, Virginia, Topographic Engineering Center, Fort Belvoir, Virginia and SPAWAR San Diego, California. This computational power is required to model futuristic sensor technology and the complexity of urban environments. For phase I the simulation generated more than two terabytes of raw data at rate of>10GB per hour. The size and distributed nature of this type of data collection posed significant challenges in developing the corresponding data-intensive applications that manage and analyze them. Building on lessons learned in developing data management tools for Urban Resolve, a next generation data management and analysis tool, called Simulation Data Grid (SDG), was developed and implemented. The design principles driving the design of SDG were

1. minimize network communication overhead (especially across SPPs) by storing data near the point of generation and only selectively propagating the data as needed, and
2. maximize the use of SPP computational resources and storage by distributing analyses across SPP sites to reduce, filter and aggregate.

The key implementation principle was to leverage existing open standards and infrastructure from Grid Computing. The developed system services interface with and were built on top of Open Grid Services Architecture standard and existing toolkits (Globus). SDG services include distributed data query/analysis, data cataloging, and data gathering/slicing/distribution. It is argued that the combination of the JESPP project and the SDG has proven to be a general-purpose tool that is extensible into a range of simulation domains.