Report No. DOEER25661
DOE Award/Contract No: DE-FG02-05ER25661
Other Identifying No.: DE-FG02-06ER25730, Amendment A000
Project Title: A Framework for Adaptable Operating and Runtime Systems

Final Technical Report for the Period 02/01/2005 – 03/31/2007

This project is conducted under the leadership and guidance of Sandia National Laboratory as part of the DOE Office of Science FAST-OS Program. It was initiated at the California Institute of Technology February 1, 2005. The Principal Investigator (PI), Dr. Thomas Sterling, accepted a position of Full Professor at the Department of Computer Science at Louisiana State University (LSU) on August 15, 2005, while retaining his position of Faculty Associate at California Institute of Technology's Center for Advanced Computing Research. To take better advantage of the resources, research staff, and students at LSU, the award was transferred by DOE to LSU where research on the FAST-OS Config-OS project continues in accord with the original proposal. This brief report summarizes the accomplishments of this project during its initial phase at the California Institute of Technology (Caltech).

During the reporting period, Caltech undertook to provide the Sandia-led project with an important test case for the composable lightweight kernel operating system being developed by the project. The Config-OS project is providing the means to implement custom lightweight operating systems to meet the distinct needs of different applications and on diverse parallel machine architectures. The dominant model of computation employed by the DOE computational science community is the "message-passing" model (more precisely: communicating sequential processes) and the Config-OS is evolving to provide strong support for this model and such widely used programming methods such as MPI. However, to fully satisfy the objectives of the project, Config-OS must be able to adapt to a diversity of driver requirements including support of a variety of execution and programming models and to do so efficiently. Towards this end, the Caltech project explored an alternative execution model to that commonly employed on scalable distributed memory systems. The objective of this work was to verify the ability and effectiveness of the Config-OS resource management system to configure a runtime environment optimized to support a broad range of execution model types and requirements.

The technical strategy of the Caltech project was to employ and extend an innovative parallel computing model, ParalleX, first developed under the previous DOE Office of Science basic research "Advanced Programming Models" project primed by Argonne National Laboratory. This early concept was devised to provide a framework to guide the development of future parallel programming languages. An initial (and admittedly rudimentary) breadboard implementation of ParalleX was accomplished in this earlier project. Although it was a sequential emulator, much insight was gained from it about future models of computation. It was from this conceptual base that the Caltech contribution to the Sandia Config-OS project was derived and ParalleX was selected as the alternative test case for the collaborative project.

ParalleX is a parallel execution model based on underlying principles of message-driven split-phase transaction multithreaded computation differing significantly from the more conventional message-passing cooperating sequential processes model in wide usage today. Its two major contributions is 1) its ability to expose and exploit diverse types, forms, and granularities of program parallelism, and 2) intrinsic system-wide latency hiding. Together, these effects can have dramatic impact on efficiency, scalability, power reduction, and performance of future high end computing systems. ParalleX supports a global name space and permits migration of parallel flow control. It further supports data-directed execution permitting effective processing of large, irregular, and time varying directed graph global data structures; an important class of algorithms employed in adaptive mesh refinement numerical problems and knowledge management symbolic computing.

The initial phase of this project conducted at Caltech achieved two important accomplishments prior to transition to LSU. The first accomplishment was to refine and extend the original ParalleX model in a number of fundamental ways to make it a better paradigm for parallel execution. The major advances include:

1. extensions and refinement of local control objects to improve fine grain synchronization,
2. representation of suspended (depleted) threads as a type of local control object,
3. clarification of intra-thread operation flow control as a static dataflow model for exposure and easy exploitation of fine grain parallelism,
4. incorporation of address translation information as part of the metadata of large directed graph data structures,
5. continuation semantics for migration of flow control through “parcel” message driven computation, and
6. application of percolation to heterogeneous (accelerators) computing.

The second major advance was the development of a ParalleX reference implementation, a completely new implementation of an emulator to support application testing of the ParalleX model. This reference implementation was developed by Maciej Brodowicz, a research staff member at the Center for Advanced Computing Research at Caltech. The ParalleX emulator was written in an Open Source version of the Common Lisp programming language and environment. This software base was selected for several reasons, in particular because it is a rich and powerful rapid prototyping environment with built in interpreter as well as compilers, interactive debugger, object-oriented semantics, sophisticated macro package, and portability across operating systems and architectures. The reference implementation is not a performance oriented systems and does not run in parallel. It has provided the basis for running applications kernels written in a variant of ParalleX Intermediate Form, or PXIF.