When DDDAS Meets Grids and HPC: A Train Wreck in the Making?

Craig C. Douglas

University of Kentucky and Yale University
douglas-craig@cs.yale.edu

In cooperation with
Gabrielle Allen and friends.

This work has been supported in part by NSF grants
EIA-0219627, ACI-0324876, and CNS-0540178 (Kentucky),

What is DDDAS

- It is a paradigm whereby application (or simulations) and measurements become a symbiotic feedback control system.
- Dynamically incorporates additional data into an executing application and lets an application dynamically steer the measurement process.
- Capabilities promise more accurate analysis and prediction, more precise controls, and more reliable outcomes.

What Benefits from DDDAS

- Applications in business, engineering and scientific processes, analysis, and design.

- Areas in manufacturing process controls, resource management, weather and climate prediction, traffic management, systems engineering, civil engineering, geological exploration, social and behavioral modeling, cognitive measurement, and bio-sensing.

How is DDDAS Done

- DDDAS creates a rich set of new challenges for applications, algorithms, systems' software, and measurement methods.
- DDDAS research typically requires
  - strong, systematic collaborations between applications domain researchers and mathematics, statistics, and computer sciences researchers
  - researchers involved in the design and implementation of measurement methods and instruments.
- Most DDDAS projects involve multidisciplinary teams of researchers.

A Few DDDAS Samples

- Weather, climate, and ocean modeling
- Flooding, typhoons or hurricanes, thunderstorms, tornados
- Fires: wildland, forest, buildings
- Contaminants: transport (water and air), identification
- Oil/gas exploration
- Traffic flow: lights and vehicular behavior

- Terrorist attack followups
- Manufacturing and full plant management
- Brain-machine interfaces
- Medical treatment (cancer treatment)
- High energy physics device control
- Mining safety would be nice

DDDAS Impact on HPC

- Scheduling
- Networking
- Data Integrity
- Allocations
- On demand everything
  - Yes, really, we need everything right now, not later, and we do not know in advance exactly what everything means

DDDAS Impact on Scheduling

- Model-model coupling across a wide area network Grid
  - Co-scheduling of both resources and networks
  - High speed versus low latency networks are an issue
- Controlling computer (e.g., a laptop, PDA, or cell phone) not necessarily part of a cluster or supercomputer center
  - Might not always on the Internet and whose IP address changes
  - Adaptive P2P scenario: Grid + separate controller
- Immediate, dynamic on demand scheduling sometimes necessary
  - Allocation change may also need to be dynamically negotiated

DDDAS Means Run It NOW

- Fast injection of application into queues
  - Fast flushing necessary
  - Must be able to write running applications to disk very fast
  - Vendors talking about modifying kernels
- Pre-injection into queues
  - Run application and go to sleep without allocating memory
  - Use small memory per processor (think Blue Gene/L or Paragon size memory usage per processsor
  - Use more processors and have small footprint per processor - saves paging time in starting up

DDDAS Impact on Data and Visualization

- Data ingestion from multiple sources
  - Authentication issues
  - Discovery
  - Networks
  - Save data that cannot be processed or *let it fall on the floor*
- Real time visualization and analysis
  - Provisioning of high speed networks
  - Flexible scheduling
  - Radically different environments may be in use at same time (Immersive or LCD screen versus PDA/cell phone)

DDDAS Impact on Computer Environment

- Information about hardware and software environment needs to be available to applications through standard APIs
  - Adapt data structures, load balancing, and computational algorithms to architecture(s)
  - Consistent dynamically linked libraries
  - Decisions for where and when to stage executables
  - Remaining queue time to prioritize work
  - Queue wait times to prioritize locations for work
  - Ability to find available hardware resources

DDDAS Impact on Performance

- Performance information needs to be available to applications through standard APIs (eg PAPI)
  - Dynamically optimize
  - Dynamically choose parameters (e.g., frequency of analysis)
- Ability to negotiate increase/decrease of allocation (time and processor count)
  - Changing workload can be caused by
    - Dynamic adaptive mesh refinement
    - Changing model (character of governing equations)

DDDAS Impact on Communication

- Internal nodes
- Third party data transfers
  - Sensor data collection
  - Controlling node is
    - Somewhere else
    - Might move on a network
- Cyber security
  - Firewalls
  - Port blocking

Conclusions

- Multiscale methodology works well and is inexpensive.
- Useful for correcting simulations' past and future predictions and leads to far fewer restarts.
- Technique can be applied to many (time dependent, nonlinear) PDE formulations used in the DDDAS field.
- HPC issues still need to be resolved and implemented seamlessly.

Request: Please send URLs for DDDAS projects to douglas-craig@cs.yale.edu for inclusion in http://www.dddas.org.