Adaptive Scheduling for Task Farming with Grid Middleware

Henri Casanova, MyungHo Kim, James S. Plank and Jack J. Dongarra

International Journal of High Performance Computing Applications 1999 13: 231
DOI: 10.1177/109434209901300306

ADAPTIVE SCHEDULING FOR TASK FARMING WITH GRID MIDDLEWARE

Henri Casanova[1]
MyungHo Kim[2]
James S. Plank[3]
Jack J. Dongarra[4]

Summary

Scheduling in metacomputing environments is an active field of research as the vision of a Computational Grid becomes more concrete. An important class of Grid applications are long-running parallel computations with large numbers of somewhat independent tasks (Monte Carlo simulations, parameter-space searches, etc.). A number of Grid middleware projects are available to implement such applications, but scheduling strategies are still open research issues. This is mainly due to the diversity of both Grid resource types and their availability patterns. The purpose of this work is to develop and validate a general adaptive scheduling algorithm for task farming applications along with a user interface that makes the algorithm accessible to domain scientists. The authors' algorithm is general in that it is not tailored to a particular Grid middleware and it requires very few assumptions concerning the nature of the resources. Their first testbed is NetSolve as it allows quick and easy development of the algorithm by isolating the developer from issues such as process control, I/O, remote software access, or fault-tolerance.

Address reprint requests to Jack J. Dongarra, Department of Computer Science, University of Tennessee, 104 Ayres Hall, Knoxville, TN 37996-1301, U.S.A.; e-mail: dongarra@cs.utk.edu.

The International Journal of High Performance Computing Applications,
Volume 13, No. 3, Fall 1999, pp. 231-240

1 Introduction

The concept of a *Computational Grid* envisioned in Foster and Kesselman (1998) has emerged to capture the vision of a network computing system that provides broad access not only to massive information resources but also to massive computational resources. Such Computational Grids will use high performance network technology to connect hardware, software, instruments, databases, and people into a seamless web that supports a new generation of computation-rich problem-solving environments for scientists and engineers. Grid resources will be ubiquitous, thereby justifying the analogy to the Power Grid.

Those features have generated interest among many domain scientists, and new classes of applications arise as being potentially *griddable*. Grid resources and their access policies are inherently very diverse, ranging from directly accessible single workstations to clusters of workstations managed by Condor (Litzkow, Livny, and Mutka, 1988), or massively parallel processor (MPP) systems with batch queuing management. Furthermore, the availability of these resources changes dynamically in a way that is close to unpredictable. Last, predicting networking behavior on the Grid is an active but still open research area. Scheduling applications in such a chaotic environment according to the end-users' need for fast response-time is not an easy task. The concept of a universal scheduling paradigm for any application at the current time is intractable, and the current trend in the scheduling research community is to focus on schedulers for broad *classes* of applications. Given the characteristics of the Grid, it is not surprising that even applications with extremely simple structures raise many challenges in terms of scheduling.

In this paper, we address applications that have simple task-parallel structures (master-slave) but require a large number of computational resources. We call such applications *task farming applications* according to the terminology introduced in Silva, Veer, and Silva (1993). Examples

[1] DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, UNIVERSITY OF CALIFORNIA AT SAN DIEGO, LA JOLLA, CALIFORNIA, U.S.A.

[2] SCHOOL OF COMPUTING, SOONGSIL UNIVERSITY, SEOUL, KOREA

[3] DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF TENNESSEE, KNOXVILLE, TENNESSEE, U.S.A.

[4] DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF TENNESSEE, KNOXVILLE, and MATHEMATICAL SCIENCE SECTION, OAK RIDGE NATIONAL LABORATORY, TENNESSEE, U.S.A.

“Our goal is not only to design a scheduling algorithm but also to provide a convenient user interface that can be used by domain scientists who have no knowledge about the Grid structure.”

of such applications include Monte Carlo simulations and parameter-space searches. Our goal is not only to design a scheduling algorithm but also to provide a convenient user interface that can be used by domain scientists who have no knowledge about the Grid structure.

Section 2 shows how some of the challenges can be addressed by using a class of Grid middleware projects as underlying operating environments, while others need to be addressed specifically with adaptive scheduling algorithms. Section 3 gives an overview of related research work and highlights the original elements of this work. Section 4 contains a brief overview of NetSolve, the Grid middleware that we used as a testbed. Sections 5 and 6 describe the implementation of the task farming interface and the implementation of the adaptive scheduling algorithm underneath that interface. Section 7 presents experimental results to validate the scheduling strategy. Section 8 concludes with future research and software design directions.

2 Motivation and Challenges for Farming

Our intent is to design and build an easily accessible computational framework for task farming applications. An obvious difficulty, then, is to isolate the users from details such as I/O, process control, connections to remote hosts, fault-tolerance, and so on. Fortunately, an emerging class of Grid middleware projects provides the necessary tools and features to transparently handle most of the low-level issues on behalf of the user. We call these middleware projects *functional metacomputing environments*. The user’s interface to the Grid is a functional remote procedure call (i.e., a call without side effects). The middleware intercepts the procedure call and treats it as a request for service. The procedure call arguments are wrapped up and sent to the Grid resources that are currently best able to service the request, and when the request has been serviced, the results are shipped back to the user, and his or her procedure call returns. The middleware is responsible for the details of managing the service on the Grid—resource selection and allocation, data movement, I/O, and fault-tolerance.

There are two main functional metacomputing environments available today. These are NetSolve (see Section 4) and Ninf (Sekiguchi et al., 1996). Building our framework on top of such architectures allows us to focus on meaningful issues like the scheduling algorithm rather than building a whole system from the ground up. Our choice of NetSolve as a testbed is motivated by the authors' experience with that system. Section 5 describes our first attempts at an application programming interface (API).

Of course, the main challenge is scheduling. Indeed, for long-running farming applications, it is to be expected that the availability and workload of resources within the server pool will change dynamically. We must therefore design and validate an adaptive scheduling algorithm (see Section 6). Furthermore, that algorithm should be general and applicable not only for a large class of applications but also for any operating environment. The algorithm is therefore designed to be portable to other metacomputing environments (Sekiguchi et al., 1996; Foster and Kesselman, forthcoming; Grimshaw et al., 1994; Litzkow, Livny, and Mutka, 1988; Abramson et al., 1997).

3 Related Work

Nimrod (Abramson et al., 1997) is targeted to computational applications based on the "exploration of a range of parameterized scenarios," which is similar to our definition of task farming. The user interfaces in Nimrod are at the moment more evolved than the API described in Section 5. However, we believe that our API will be a building block for high-level interfaces (see Section 8). The current version of Nimrod (or Clustor, the commercial version available from http://www.activetools.com) does not use any metacomputing infrastructure project, whereas our task farming framework is built on top of Grid middleware. However, a recent effort, Nimrod/G (Abramson and Giddy, 1997), plans to build Nimrod directly on top of Globus (Foster and Kesselman, forthcoming). We believe that a project like NetSolve (or Ninf) is a better choice for this research work. First, NetSolve is freely available. Second, NetSolve provides a very simple interface, letting us focus on scheduling algorithms rather than Grid infrastructure details. Third, NetSolve can and probably will be implemented on top of most Globus services and will then leverage the Grid infrastructure without modifications of our scheduling algorithms. Another distinction between this work and Nimrod is that the latter does not contain adaptive algorithms for scheduling like the one described in Section 6. In fact, it is not inconceivable that the algorithms eventually produced by this work could be incorporated seamlessly into Nimrod.

Calypso (Baratloo , Dasgupta, and Kedem, 1995) is a programming environment for a loose collection of distributed resources. It is based on C++ and shared memory but exploits task-based parallelism of relatively independent jobs. It has an eager scheduling algorithm and, like the functional metacomputing environments described in this paper, uses the idempotence of the tasks to enable a replication-based fault-tolerance.

A system implemented on top of the Helios OS that allows users to program master-slave programs using a "Farming" API is described in Silva, Veer, and Silva (1993) and Silva et al. (1995). Like in Calypso, the idempotence of tasks is used to achieve fault-tolerance. They do not focus on scheduling.

The AppLeS project (Berman et al., 1996; Berman and Wolski, 1997) develops metacomputing scheduling agents for broad classes of computational applications. Part of the effort targets scheduling master-slave applications (Berman, Wolski, and Shao, 1998) (task-farming applications with our terminology). A collaboration between the NetSolve and the AppLeS team has been initiated, and integration of AppLeS technology, the Network Weather Service (NWS) (Wolski, 1996), NetSolve-like systems, and the results in this document is under way.

As mentioned earlier, numerous ongoing projects are trying to establish the foundations of the Computational Grid envisioned in Foster and Kesselman (1998). Ninf (Sekiguchi et al., 1996) is similar to NetSolve in that it is targeted to domain scientists. Like NetSolve, Ninf provides simple computational services, and the development teams are collaborating to make the two systems interoperate and standardize the basic protocols. At a lower level are Globus (Foster and Kesselman, forthcoming) and Legion (Grimshaw et al., 1994), which aim at providing basic infrastructure for the Grid. Condor (Litzkow, Livny, and Mutka, 1988; Litzkow and Livny, 1990) defines and implements a powerful model for Grid components by allowing the idle cycles of networks of workstation to be harvested for the benefit of Grid users without penalizing local users.

4 Brief Overview of NetSolve

The NetSolve project is under development at the University of Tennessee and the Oak Ridge National Laboratory. Its original goal is to alleviate the difficulties that domain scientists usually encounter when trying to locate/install/use numerical software, especially on multiple platforms. With NetSolve, the user does not need to be concerned with the location/type of the hardware resources being used or with the software installation. Furthermore, NetSolve provides transparent fault-tolerance mechanisms and implements scheduling algorithms to minimize overall response time. As seen in Figure 1, NetSolve has a three-tiered design in that a *client* consults an

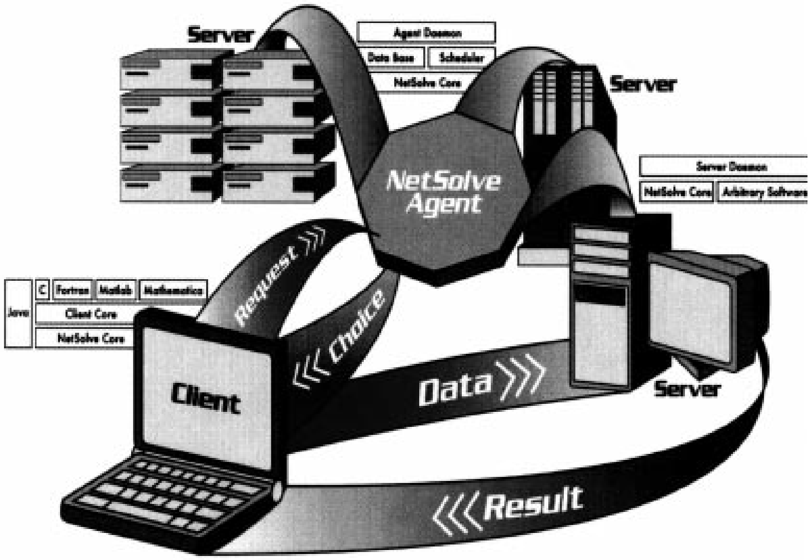

Fig. 1 The NetSolve System

agent prior to sending requests to a *server.* Let us give basic concepts about those three components as well as information about the current status of the project.

The NetSolve Server. A NetSolve server can be started on any hardware resource (single workstation, cluster of workstations, MPP). It can then provide access to arbitrary software installed on that resource (NetSolve provides mechanisms to integrate any software component into a server so that it may become available to NetSolve clients; Casanova and Dongarra, 1998a).

The NetSolve Agent. The NetSolve agent is the key to the computation-resource mapping decisions as it maintains a database about the statuses and capabilities of servers. It uses that database to make scheduling decisions for incoming user requests. The agent is also the primary participant in the fault-tolerance mechanisms. Note that there can be multiple instances of the NetSolve agent to manage a confederation of servers.

The NetSolve Client. The user can submit (possibly simultaneous) requests to the system and retrieve results with one of the provided interfaces (C, Fortran, Matlab [see Math Works, 1992], Mathematica [see Wolfram, 1996], Java APIs, or Java GUI).

Current Status of NetSolve. At this time, a preversion of NetSolve 1.2, containing full-fledged software for all UNIX flavors, Win32 C, and Matlab APIs, can be downloaded from the homepage at

http://www.cs.utk.edu/netsolve.

The NetSolve users' guide (Casanova, Dongarra, and Seymour, 1996) contains general purpose information and examples. Details about the NetSolve agent can be found in Casanova and Dongarra (1997). Recent developments and applications of NetSolve are described in Casanova and Dongarra (1998b). Last, technical details about the current NetSolve implementation are to be found in (Casanova and Dongarra, 1998c).

5 Task Farming API

5.1 BASICS

In this work, we assume that a functional metacomputing environment is available (see Section 2). That environment provides an API that contains two functions: (i) `submit()` to send a request asynchronously for computation and (ii) `poll()` to poll asynchronously for the completion of a request. Polling returns immediately with the status of the request. If the computation is complete, the result is returned as well. The NetSolve and Ninf APIs satisfy these requirements. In addition, the environment provides access to preinstalled software and hardware resources. The user just provides input data and a way to identify which software should be used to process that data. Again, both NetSolve and Ninf comply.

A farming job is one composed of a large number of independent requests that may be serviced simultaneously. This is sometimes referred to as the "bag-of-tasks" model (Bakken and Schilchting, 1995; Gelernter and Kaminsky, 1992). Farming jobs fall into the class of "embarrassingly parallel" programs, for which it is very clear how to partition the jobs for parallel programming environments. Many important classes of problems, such as Monte Carlo simulations (e.g., Stiles et al., 1998) and parameter-space searches (e.g., Abramson et al., 1997) fall into this category.

Without a farming API, the user is responsible for managing the requests himself or herself. One possibility would be to submit all the desired requests at once and let the system schedule them. However, we have seen that scheduling on the Grid is a challenging issue, and as a result, the available Grid middleware projects implement only minimal scheduling capabilities that do not optimize even this simple class of parallel programs. A second possibility is for the user to manually manage a ready queue by having at most *n* requests submitted to the system at any point in time. This solution seems more reasonable; however, the optimal value of *n* depends on Grid resource availability, which is beyond the user's control and is dynamic.

5.2 API

It is difficult to design an API that is both convenient for the end-user and sophisticated enough to handle many real applications. Our farming API contains one function, `farm()`, with which the user specifies all data for all the computation tasks. The main idea is to replace multiple calls to `submit()` by one call to `farm()` whose arguments are lists of arguments to `submit()`. In this first implementation, we assume that arguments to `submit()` are either integers or pointers (which is consistent with the NetSolve specification). Extending the call to support other argument types would be trivial. The first argument to `farm()` specifies the number of requests by declaring an induction variable and defining its range. The syntax if "`i = %d,%d`" (see example below). The second argument is the identifier for the computational functionality in the metacomputing environment (a string with Ninf and NetSolve). Then follow a (variable) number of argument lists. Our implementation provides three functions that need to be called to generate such lists: (i) `expr()` allows an argument to computation *i* to be an integer computed as an arithmetic expression containing *i*; (ii) `int_array()` allows an integer argument to computation *i* to be an element of an integer array indexed by the value of an arithmetic expression containing *i*; (iii) `ptr_array()` is similar to `int_array()` but handles pointer arguments. Arithmetic expressions are specified with Bourne Shell syntax (accessing the value of *i* with '`$i`').

Let us show an example assuming that the underlying metacomputing environment provides a computational function called "foo." The code

```
double x[10],y[30],z[10];
submit("foo",2,x,10);
submit("foo",4,y,30);
submit("foo",6,z,10);
```

makes three requests to run the "`foo`" software with the following sets of arguments: (`2,x,10`), (`4,y,30`), and (`6,z,10`). Note that `x`, `y`, and `z` will hold the results of the NetSolve call. With farming, these calls are replaced by

```
void *ptrs[3];
int  *ints[3];

ptrs[0] = x;  ptrs[1] = y;  ptrs[2] = z;
ints[0] = 10; ints[1] = 30; ints[2] = 10;

farm("i=0,2","foo",expr("2*($i+1)"),
   ptr_array(ptrs,"$i"),int_array
   (ints,"$i"));
```

We expect to use this API as a basis for more evolved interfaces (e.g., graphical or Shell-based). So far, we have used the API directly to implement basic example computations (2D block-cyclic matrix-multiply, Mandelbrot set computation) and to build a Shell interface to MCell (see Section 7). Section 8 describes how we plan to generalize this work to automatically generate high-level interfaces.

"Our farming algorithm manages a ready queue and adapts to the underlying metacomputing environment . . ."

```
$n$ = initial guess on the queue size;
$\alpha$ = scheduling factor;
$\delta = 1$;
while (tasks remaining) {
    while (number of pending tasks < $n$) {
        submit();
    }
    foreach (pending task) {
        poll();
    }
    if ($n$ – number of pending tasks $\geq n \times \alpha$) {
        if (average task response time has improved) {
            $n = n + \delta$;
      $\delta = \delta + 1$;
        }
        else {
            $n = n - \delta$;
            $\delta = 1$;
        }
    }
}
```

Fig. 2 Adaptive scheduling algorithm

6 Scheduling Strategy

6.1 THE SCHEDULING ALGORITHM

The main idea behind the scheduling algorithm has already been presented in Section 5.1: managing a ready queue. We mentioned that the user had no elements on which to base the choice for n, the size of the ready queue. Our farming algorithm manages a ready queue and adapts to the underlying metacomputing environment by modifying the value of n dynamically according to constant computation throughput measurement. The algorithm really sees the environment as an opaque entity that gives varying responses (request response times) to repeated occurrences of the same event (the sending of a request).

Let us go through the algorithm shown in Figure 2. First, the algorithm chooses the initial value of n. That choice can be arbitrary, but it may benefit from additional information provided by the underlying metacomputing environment. NetSolve provides a way to query the agent about the number of available servers for a given computation, and that number is the initial guess for n in this first implementation. Second, the algorithm sets the *scheduling factor* α, which takes values in (0,1) and determines the behavior of the algorithm. Indeed, the value of n may be changed only when more than n tasks completed during one iteration of the outermost while loop. A value of $\alpha = 1$ causes the algorithm to be extremely conservative (only when all n requests are completed instantly may the value of n be changed). The smaller the α, the more often will the algorithm try to modify n. The algorithm keeps a running history of the average request response times for all requests in the queue. That history is used to detect improvements or deterioration in performance and modify the value of n accordingly.

This algorithm is rather straightforward at the moment, but it will undoubtedly be improved after more experiments have been conducted. However, early experimental results shown in Section 7 are encouraging.

6.2 CURRENT IMPLEMENTATION

In our testbed implementation of farming for NetSolve, we implement `farm()` as an additional layer on top of the traditional NetSolve API, exactly as detailed in Section 5.2. A similar implementation would be valid for a system like Ninf. In other metacomputing environments, placing the scheduling algorithm within the client library might not be feasible, in which case the algorithm needs to be implemented in other parts of the system (central scheduler, client proxy, etc.). However, the algorithm is designed to rest on top of the metacomputing system, rather than to be merged with the internals of the system.

6.3 POSSIBLE EXTENSIONS

The NetSolve farming interface is very general, and we believe that it can serve as a low-level building-block for deploying various classes of applications. However, this generality leads to shortcomings. The embedded scheduler cannot take advantage of application-specific features, such as exploitable data patterns. Real applications are likely to manipulate very large amounts of data, and it may be possible for the scheduler to make decisions based on I/O requirements. For instance, one can imagine that a subset of the tasks to farm makes use of one or more constant input data. This is a frequent situation in MCell (see Section 7.1), for example. Such input data could then be *shared* (via files, for instance) by multiple resources, as opposed to being replicated across all the resources. Another possibility would be for the farming application to contain simple data dependencies between tasks. In that case, our framework could detect those dependencies and schedule the computations accordingly. Another shortcoming of the farming interface that is a direct cause of its generality is that the call to `farm()` is completely atomic. This is an advantage from the point of view of ease-of-use, but it prevents such things as visualization of results as they become available, for instance. Once again, such a feature would be desirable for MCell. Section 8 lays the ground for research in these directions, and work is under way in the context of MCell.

7 Preliminary Experimental Results

7.1 MCELL

MCell (Stiles et al., 1998; Stiles et al., 1996) is a general Monte Carlo simulator of cellular microphysiology. MCell uses Monte Carlo diffusion and chemical reaction algorithms in 3D to simulate the complex biochemical interactions of molecules inside and outside of living cells. MCell is a collaborative effort between the Terry Sejnowski lab at the Salk Institute and the Miriam Salpeter lab at Cornell University. Like any Monte Carlo simulation, MCell must run large numbers of identical, independent simulations for different values of its random number generator seed. It therefore qualifies as a task farming application and was our first motivation to develop a farming API along with a scheduling algorithm.

As mentioned earlier, we developed for MCell a Shell-based interface on top of the C farming API. This interface takes as input a user-written *script* and automatically generates the call to `farm()`. The script is very intuitive as it follows the MCell command-line syntax by just adding the possibility for *ranges* of values as opposed to fixed values. For instance, instead of calling MCell

```
mcell foo1 1
mcell foo1 2
….
mcell foo1 100
```

it is possible to call MCell

```
mcell foo1 [1-100]
```

which is simpler, uses Grid computational resources from NetSolve, and ensures good scheduling with the use of the algorithm described in Section 6.

7.2 RESULTS

The results presented in this section were obtained by using a NetSolve system spanning 5 to 25 servers on a network of Sun workstations (Sparc ULTRA 1) interconnected via 100Mb Ethernet. The farming application uses MCell to compute the shape of the parameter space, which describes the possible modes of operation for the process of synaptic transmission at the vertebrate neuromuscular junction. Since MCell's results include the true stochastic noise in the system, the signal must be averaged at each parameter space point. This is done by running each point 10 times with 10 different values of the random number generator seed. In this example, three separate 3D parameter spaces are sampled, each parameter space is of dimension $3 \times 3 \times 3$. The number of tasks to farm is therefore $3 \times 3 \times 3 \times 3 \times 10 = 810$, and each task generates 10 output files.

These preliminary experiments were run on a dedicated network. However, we simulated a dynamically changing resource pool by linearly increasing and decreasing the number of available NetSolve computational servers. Results are shown in Table 1 for our adaptive scheduling, a fixed queue size of $n = 25$, and a fixed queue size of $n = 5$.

Table 1
Preliminary Experimental Results

Scheduling	Time	Resource Availability (%)	Relative Performance (%)
Adaptive	3982 s	64	100
n = 25	4518 s	62	85
n = 5	10,214 s	63	38

> ***"In this article, we have motivated the need for schedulers tailored to broad classes of applications running on the Computational Grid."***

The resource availability measures the fraction of servers available during one run of the experiment. As this number changes throughout time, the availability is defined as the sum of the number of servers available over all time steps (10 seconds). We compare scheduling strategies by measuring *relative performance*, which we define as a ratio of *adjusted elapsed times*, taking the adaptive scheduling as a reference. Adjusted elapsed times are computed by assuming a 100% availability and scaling the real elapsed times accordingly. One can see that the adaptive strategy performs 15% better than the strategy with *n* = 25. Of course, the strategy with *n* = 5 performs very poorly since it does not take advantage of all the available resources.

These first results are encouraging but not as satisfactory as expected. This is due to the implementation of NetSolve and the way the experiment was set up. Indeed, NetSolve computational tasks are not interrupted when a NetSolve server is terminated. Terminating a server only means that no further requests will be answered but that pending requests are allowed to terminate. Thus, this experiment does not reflect the worst-case scenario of machines being shut down causing all processes to terminate. We expect our adaptive strategy to perform even better in an environment where tasks are terminated prematurely and need to be restarted from scratch on remaining available resources. Due to time constraints, this article does not contain results to corroborate this assumption, but experiments are under way.

8 Conclusion and Future Work

In this paper, we have motivated the need for schedulers tailored to broad classes of applications running on the Computational Grid. The extreme diversity of Grid resource types, availabilities, and access policies makes the design of schedulers a difficult task. Our approach is to build on existing and available metacomputing environments to access the Grid as easily as possible and to implement an interface and scheduling algorithm for task farming applications. An adaptive scheduling algorithm was described in Section 6. That algorithm is independent from the internal details of the Grid and of the metacomputing environment of choice. We chose NetSolve as a testbed for early experiments with the MCell application. The effectiveness of our scheduler is validated by preliminary experimental results in Section 7. Thanks to our framework for farming, a domain scientist can easily submit large computations to the Grid in a convenient manner and have an efficient adaptive scheduler manage execution on his or her behalf.

There are many ways in which this work can be further extended. We already mentioned in Section 6.3 that it is possible to use the farming API to detect data dependencies or shared input data between requests. The adaptive scheduling algorithm could be augmented to take into account such patterns. A possibility is for the farming interface to take additional arguments that describe domain-specific features and that may activate more sophisticated scheduling strategies if any. A first approach would be to consider only input or output coming from files (which is applicable to MCell and other applications) and partition the request space such as to minimize the number of file transfers and copies. This will require that the underlying metacomputing environment provide a feature to describe such dependencies. Work is being done in synergy with the NetSolve project to take into account data locality, and the farming interface will undoubtedly take advantage of these developments (Beck et al., forthcoming). This will be fertile ground for scheduling and data logistic research. The scheduling algorithm can also be modified to incorporate more sophisticated techniques. For instance, if the metacomputing environment provides an API to access more details about the status of available resources, it might be the case that *n*, the size of the ready queue, can be tuned effectively. The danger, however, is to lose portability as the requirements for the metacomputing environment (see Section 5.1) would be more stringent. Experiments will be conducted in order to investigate whether such requirements can be used to significantly improve scheduling.

The farming API can be enhanced so that certain tasks may be performed upon submitting each request and receiving each result. For instance, the user may want to visualize the data as they are coming back, as opposed to waiting for completion of all the requests. This is not possible at the moment as the call to `farm()` is atomic and does not provide control over each individual request. A possibility would be to pass pointers to user-defined functions for `farm()` and execute them for events of interest (e.g., visualization for each reception of a result). Such functions could take arbitrary arguments for the sake of versatility. Some of the available metacomputing environments provide attractive interactive interface to which a farming call could be contributed. Examples include Matlab (NetSolve) and Mathematica (NetSolve, Ninf). To make our task-farming framework easily accessible to a growing number of domain scientists, we need to develop ways to use the C farming API as a basis for more usable high-level interfaces. Steps in that direction have already been taken with the Shell-interface for MCell (see Section 7.1). It would be rather straightforward to design or use an existing specification language to describe specific farming applications and automatically generate custom Shell-based graphical interfaces like the ones in Abramson et al. (1997).

ACKNOWLEDGMENTS

This material is based on work supported by the National Science Foundation under grant CCR-9703390, by NASA/NCSA Project Number 790NAS-1085A, under Subaward Agreement #790, and by the National Science Foundation Science and Technology Center Cooperative Agreement No. CCR-8809615.

BIOGRAPHIES

Henri Casanova is a project scientist at the University of California at San Diego. His research interests include all areas of metacomputing, and in particular theoretical models and simulation techniques for predicting and forecasting the performance of globally or locally distributed applications, in a view to the efficient scheduling of these applications in a computational Grid environment. He received his B.S. in computer science and applied mathematics from the Ecole Nationale Supérieure d'Electrotechnique, d'Informatique et d'Hydraulique de Toulouse (ENSEEIHT), his M.S. in parallel architectures and applied mathematics from the University Paul Sabatier, Toulouse, and his Ph.D. in computer science from the University of Tennessee, Knoxville.

MyungHo Kim received a B.A. in computer science in 1989 from SoongSil University and M.S. and Ph.D. degrees in computer science from POSTECH in 1991 and 1995, respectively. Since September 1995, he has been an associate professor in the School of Computing at SoongSil University, and since July 1998 he has been a visiting scholar in the computer science department at the University of Tennessee. From November 1994 to August 1995, he was a senior researcher in the computer technology division at ETRI. His research interests are in parallel and distributed computing, parallel algorithm, and parallel software tools.

James S. Plank received his B.S. from Yale University in 1988 and his Ph.D. from Princeton University in 1993. He is currently an associate professor in the computer science department at the University of Tennessee. His research interests are in fault-tolerance, network computing, and operating systems.

Jack J. Dongarra holds a joint appointment as Distinguished Professor of Computer Science in the computer science department at the University of Tennessee (UT) and as Distinguished Scientist in the Mathematical Sciences Section at Oak Ridge National Laboratory (ORNL) under the UT/ORNL

Science Alliance Program. He specializes in numerical algorithms in linear algebra, parallel computing, use of advanced-computer architectures, programming methodology, and tools for parallel computers. Other current research involves the development, testing, and documentation of high quality mathematical software. He was involved in the design and implementation of the software packages EISPACK, LINPACK, the BLAS, LAPACK, ScaLAPACK, Netlib, PVM, MPI, the National High-Performance Software Exchange, NetSolve, and ATLAS and is currently involved in the design of algorithms and techniques for high performance computer architectures.

REFERENCES

Abramson, D., I. Foster, J. Giddy, A. Lewis, R. Sosic, and R. Sutherst. 1997. The Nimrod Computational Workbench: A case study in desktop metacomputing. Paper presented at Proceedings of the 20th Autralasian Computer Science Conference, February, Sidney, Australia.

Abramson, D., and J. Giddy. 1997. Scheduling large parametric modelling experiments on a distributed meta-computer. Paper presented at PCW'97, September.

Bakken, D. E., and R. D. Schilchting. 1995. Supporting fault-tolerant parallel programming in Linda. *IEEE Transactions on Parallel and Distributed Systems* 6 (3): 287-302.

Baratloo, A., P. Dasgupta, and Z. Kedem. 1995. Calypso: A novel software system for fault-tolerant parallel processing on distributed platforms. Paper presented at the 4th IEEE International Symposium on High Performance Distributed Computing, August, Pentagon City, VA.

Berman, F., and R. Wolski. 1997. The AppLeS Project: A status report. Paper presented at Proceedings of the 8th NEC Research Symposium, May, Berlin, Germany.

Berman, F., R. Wolski, S. Figueira, J. Schopf, and G. Shao. 1996. Application-level scheduling on distributed heterogeneous networks. Paper presented at Proceedings of Supercomputing'96, November, Pittsburgh, PA.

Berman, F., R. Wolski, and G. Shao. 1998. Performance effects of scheduling strategies for master/slave distributed applications. TR-CS98-598. University of California, San Diego.

Casanova, H., and J. Dongarra. 1997. NetSolve: A network server for solving computational science problems. *The International Journal of Supercomputer Applications and High Performance Computing* 11 (3): 212-23.

Casanova, H., and J. Dongarra. 1998a. Providing uniform dynamic access to numerical software. In *IMA volumes in mathematics and its applications, algorithms for parallel processing*. Vol. 105. Edited by M. Heath, A. Ranade, and R. Schrieber, 345-55. New York: Springer-Verlag.

Casanova, H., and J. Dongarra. 1998b. NetSolve's network enabled server: Examples and applications. *IEEE Computational Science & Engineering* 5 (3): 57-67.

Casanova, H., and J. Dongarra. 1998c. NetSolve version 1.2: Design and implementation, Department of Computer Science, University of Tennessee, Knoxville.

Casanova, H., J. Dongarra, and K. Seymour. 1996. Client user's guide to NetSolve. CS-96-343. Department of Computer Science, University of Tennessee, Knoxville.

Foster, Ian, and Carl Kesselman. 1997. Globus: A metacomputing infrastructure toolkit. *International Journal of Supercomputer Applications* 11 (2): 115-28.

Foster, Ian, and Carl Kesselman. 1998. *The Grid: Blueprint for a new computing infrastructure*. San Francisco: Morgan Kaufmann.

Gelernter, D., and D. Kaminsky. 1992. Supercomputing out of recycled garbage: Preliminary experience with Piranha. Paper presented at International Conference on Supercomputing, June, Washington, DC.

Grimshaw, A., W. Wulf, J. French, A. Weaver, P. Reynolds Jr. 1994. A synopsis of the Legion Project. CS-94-20. Department of Computer Science, University of Virginia, Charlottesville.

Litzkow, M., and M. Livny. 1990. Experience with the Condor Distributed Batch System. Paper presented at Proceedings of IEEE Workshop on Experimental Distributed Systems, October, Department of Computer Science, University of Wisconsin, Madison.

Litzkow, M., M. Livny, and M. W. Mutka. 1988. Condor—A hunter of idle workstations. Paper presented at Proceedings of the 8th International Conference of Distributed Computing Systems, June, Department of Computer Science, University of Wisconsin, Madison.

The Math Works. 1992. MATLAB reference guide. Nitick, MA: Math Works.

Sekiguchi, S., M. Sato, H. Nakada, S. Matsuoka, and U. Nagashima. 1996. Ninf: Network based Information Library for Globally High Performance Computing. Paper presented at Proceedings of Parallel Object-Oriented Methods and Applications (POOMA), February, Santa Fe, NM.

Silva, L., B. Veer, and J. Silva. 1993. How to get a fault-tolerant farm. *World Transputer Congress*, September, pp. 923-38.

Silva, L. M., J. G. Silva, S. Chapple, and L. Clarke. 1995. Portable checkpointing and recovery. Paper presented at Proceedings of the HPDC-4, High Performance Distributed Computing, August, Washington, DC.

Stiles, J. R., T. M. Bartol, E. E. Salpeter, and M. M. Salpeter. 1998. Monte Carlo simulation of neuromuscular transmitter release using MCell, a general simulator of cellular physiological processes. In *Computational neuroscience*, ed. J. M. Bower, 279-84. New York: Plenum.

Stiles, J. R., D. Van Helden, T. M. Bartol, E. E. Salpeter, and M. M. Salpeter. 1996. Miniature end-plate current rise times 100 microseconds from improved dual recordings can be modeled with passive acetylcholine diffusion form a synaptic vesicle. *Proc. Natl. Acad. Sci. U.S.A.* 93:5745-52.

Wolfram, S. 1996. *The mathematica book*. 3d ed. Cambridge, UK: Wolfram Median and Cambridge University Press.

Wolski, R.. 1996. Dynamically forecasting network performance using the Network Weather Service. TR-CS96-494. University of California, San Diego.