Consistency Management for Data Grid in OptorSim Simulator

Ghalem Belalem
Dept. of Computer Science, Faculty of Sciences,
University of Oran - Es Senia, Oran, Algeria
BP. 1524, El M'Naouer, 31000 Oran, Algeria
ghalem1dz@yahoo.fr

Yahya Slimani
Dept. of Computer Science, Faculty of Sciences of Tunis,
1060 Tunis, Tunisia
yahya.slimani@fst.rnu.tn

Abstract

One of the principal motivations to use the grids computing and data grids comes from the applications using of large sets from data, for example, in High-Energy physics or Life Science to improve the total output of the software environments used to carry these applications on the grids, data replication are deposited on various selected sites. In the field of the grids the majority of the strategies of replication of the data and scheduling of the jobs were tested by simulation. Several simulators of grids were born. One of the most simulators interesting for our study is the OptorSim tool. In this paper, we present an extension of the OptorSim simulator by a consistency management module of the replicas in the Data Grids. This extension corresponds to a hybrid approach of consistency, it inspired by the pessimistic and optimistic approaches of consistency. This suggested approach has two vocations, in the first time, it makes it possible to reduce the response times compared with the completely pessimistic approach, in the second time, it gives a good quality of service compared with the optimistic approach.

1 Introduction

The collection of the great data bases reached rather significant sizes measured in Petabytes. The communities of the researchers who must reach and analyze these data are often significant, in the same way for computing and data resources. This combination of data bases of big size, the users and the resources with an intensive calculation requires a new infrastructure of management. A great collaboration of the scientists can produce many requests, each one implying of a compute and the access to the great sets of data. The reliable and effective execution of these requests requires a careful management of the data, wide area networks of high flow and other advanced techniques which maximize the use of storage collectively, of the management of the networks and the data-processing resources. The response to these changes is to pass to a model of data processing distributed making it possible to fully exploit the resources and the capacities offered. This environment will offer a service and an access uniform and economically viable to the resources of infrastructure. This evolution is known under the name of *grids computing and data grids* [11]. One of the objectives of data grids is the effective management of data replication in a transparent and reliable way. Several simulators were proposed to study the behavior and the evolution of this type of infrastructure, among them, we can quote: Brigs [18], SimGrid [8], GridSim [6], ChicSim [15], EdgSim [1], MicroGrid [17], GangSim [10], OptorSim [5]. This work consists in extending the tool OptorSim, a simulator of data grids intended for management of the replication, by a manager of consistency of which its main goal makes, it possible to maintain the consistency of replicas in data grid. Our article will be structured as follows: in section 2, we will present the tool OptorSim and its characteristics, section 3 will be intended for the approaches of replication (pessimistic or optimistic) and their management, section 4 will be reserved for the description of our hybrid approach proposed. Section 5 will describe some metrics which we consider significant in interpretations of the results. Section 6 will be reserved for some resulting from experiments preliminary by using the OptorSim simulator. Finally we will finish by some directions for future work.

2007 International Conference on Multimedia and Ubiquitous Engineering(MUE'07)
0-7695-2777-9/07 $20.00 © 2007 IEEE

2 OptorSim simulator

OptorSim [4, 5, 7] is a package of simulation written in Java which is used to modeling the interactions of the individual components of data grid and its basic architecture of OptorSim simulator is presented in figure 1. Its design derives directly from the architecture of the Data Grid project.

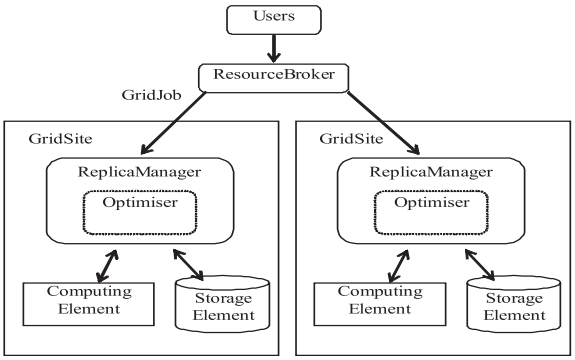

Figure 1. Basic architecture of OptorSim

The principal motivation of OptorSim was the need of environments of simulation for applications of treatment of great whole of data in data grid. One of the principal concepts in data grids is the data replication, the objective of creation and the management of data replication in various geographical places are to optimize the cost of access to the data and to study the stability and the transitory behavior of the methods of optimization of replicas. The grid consists of sites offering each resource to the jobs to be executed. The resources computing are named CE (Computing Element) and the resources of storage are named SE (Storage Element). CE execute the jobs which use the data of the files stored in the SE. A Resource Broker control the scheduling of the jobs in various CE. For goal to simulate our protocol suggested of management of coherence in the environments of the grids, the choice was related to the OptorSim simulator for the principal reasons: OptorSim is a software Open Source written in Java, it was conceived to study and test the dynamic strategies of replication [9, 12, 14]. The simulation of the grid resources is the base of simulation of the grid environment. These resources include mainly the computing resources, the storage resources which are connected in a network environment. The simulation of grid resources can comprise mainly the components:

- Simulation of computing resources: have the capacities of calculus;
- Simulation of storage resources:present the property of the simulation of storage spaces;
- Simulation of network environment: comprise mainly three parameters: Bandwidth of network, analyzes network and the detention of network;

Finally, OptorSim is a Grid simulator designed to test dynamic replication strategies used in optimizing the efficiency of a Grid. OptorSim takes a grid configuration and a replica optimizer algorithm as input and then runs a number of grid jobs using the given configuration. It also allows a user to visualize the performance of the algorithm.

3 Consistency management approaches

The Consistency [3, 9] is a relation which defines the degree of similarity between copies of a distributed entity. In the ideal case, this relation characterizes copies which have identical behaviors. In the real cases, where the copies evolve in a different way, consistency defines the limits of divergence authorized between these copies [13, 16]. We need a consistency protocol which ensures the execution of the operations of users, the mutual consistency of copies in accordance with a behavior defined by a model of coherence. The consistency protocol gives an ideal view as if there is only one user and only one copy of the data in the system. Replica consistency management can be achieved either synchronously using the so-called pessimistic algorithms, or asynchronously deploying optimistic ones. Fundamental tussle between pessimistic and optimistic approach is that of scalability and security. The execution of pessimistic consistency assures that any change in one replica is atomically propagated to all other replicas. Therefore, there is an inherent guarantee that all replicas will have the same data at all time, making this approach indispensable in the mission of critical and sensitive applications like the distributed banking application .On the other hand, the optimistic approach is employs for applications, which evolves quickly, in mobile environments and system weakly coupled. So that we can say that, the pessimistic approach is interested in consistency more than availability, while the optimistic approach supports the availability more than the consistency.

3.1 Pessimistic approach

The pessimistic approach prohibits any access to a replica unless it is provably up to date [16]. This make users believe that they have only one consistent copy. The main advantage of this approach is that all replicas converge at the same time, and guarantees high consistency of data. This approach is well adapted to small and middle scale systems, but becomes very complex when it is applied for large scale systems . Two major drawbacks of this approach have to be noted: To execute a pessimistic algorithm one is obliged to execute the following actions:

- Block the replica, deliver messages using total ordering or implement voting protocol. In more active partition and synchronous operations from all the nodes is required. That is what makes the model well suited only for LAN based environments where the network is stable because message delivery is guaranteed and network partitions are non-existent;

- It is very badly adapted to uncertain and unsteady environments, such as mobile systems, and data grids with high rate of changes.

3.2 Optimistic approach

This also means the optimistic strategy allow users to reach any copy for the reading or the writing operations, even when there are breakdowns of network or when some copies are unavailable [16]. This also means that the approach can lead to replica inconsistency. On the other hand, the approach requires a follow-up phase to detect and then correct divergences between replicas by converging them toward a coherent state. Although this approach does not guarantee a high consistency with respect to the pessimistic one. One can also indicate some disadvantages of the optimistic approach like:

- The states of copies can be temporarily mutually contradictory;

- An update can be applied to one copy without being synchronically applied to other copies, and there will can be even a substantial time since the application of an update in a copy until the propagation of the update to other copies. The concurrent updates with the various copies can present conflicts. For example, in a distributed system of reservation of air line which uses the optimistic strategy of consistency, two copies can accept a reservation for the same seat[16].

4 Model proposed

4.1 Architecture of model proposed

Our proposed approach combines the optimistic and pessimistic approaches to ensure replica consistency of data in a grid [3]. This approach uses a hierarchical model of a grid where the replicas of a data are located. This hierarchical model is two-based and it is composed by only two levels *(see figure 2)*. In our work, we consider a grid as a collection of distributed collection of Computing Elements (CE's) and Storage Elements (SE's). These elements are linked together through a network to form a Site or a Cluster.

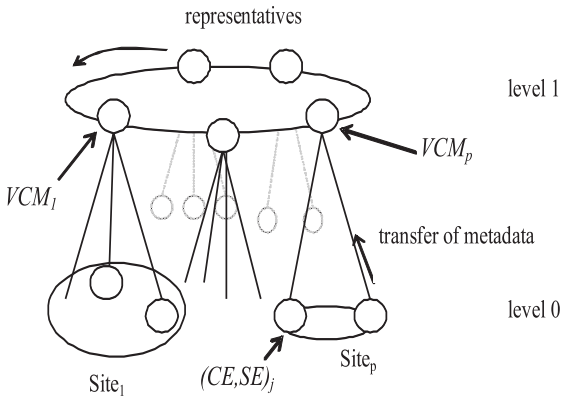

Figure 2. Model proposed architecture

1. Level 0 : in this level we find sites that compose a grid. Each site contains a set of Computing Elements (CE's) and Storage Elements (SE's). Replicated data are stored on SE's and accessed from CE's via reading or writing operations. Each replica attached to additional information is called metadata (TimeStamp, indices, versions,...).

2. Level 1 : This level is also called inter-sites consistency, it is responsible for global consistency in the data grid, every site cooperates with the other sites via a an elementary constituent. for this goal we define in this level and for every $Site_i$ a virtual representative VCM_i (Virtual Consistency Manager) a virtual representative who communicates with its homologue VCM_j of another $Site_j$ this communication is implemented due to a getaway, and based on the transfer of metadata. For them $k-1$ remaining nodes of every site are responsible for receiving all the requests emanating from users, these requests can be of writing or of reading. We note, every node is a couple of (CE, SE) where the CE is a Computing Element, who define the functionalities of a each VCM (see Figure 2).

Our model proposed can have several interesting characteristics, we can quote for example:

- Simplicity: the model is simple, by the fact that it has only two levels;

- Topology of the sites can be variable (ring, star, ...);

- The consistency strategies of the sites can be variable (Rowa, quorum, ...);

- Reduction of the aspect of communication;

- Scalability.

4.2 Process of hybrid protocol

Our approach is articulated on two modules: the first is concerned by local consistency and the second is concerned by global consistency. These modules cooperate to maintain the consistency of all the system. They interact with the components of the OptorSim simulator, as it is shown in the figure 3.

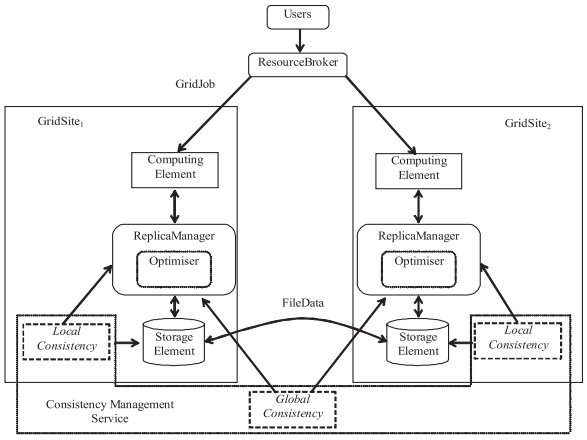

Figure 3. Structure of OptorSim extended by the consistency service

4.2.1 What is the local consistency?

Local consistency is also called consistency intra-site. Its principal objective is to ensure consistency in a continuous way between the various replicas of the same data inside a site, which corresponds to make converge the replicas towards a relative replica for a site and it is founded on the optimistic approach of replication. It is started in an alternative way with global consistency. Its principal stapes are:

i) Replication strategy: This task defines the replication policy to use at the layer of a given site. The policy can different from one site to another. Thus, it will be possible to apply customized policies based on techniques such as single master, multi-masters, quorum, etc. For more details about the existing techniques;

ii) Treatment of the request: With the reception of a request, subjected by a client towards a given site, it is immediately treated by the CE according to the strategy of replication of the site receiving and sent to the customer. For a request of writing, information of the metadata of this replica will be updated (version, timestamp, ...);

iii) Propagation of the update: In the event of a request of the writing type, a propagation of the updates is started for sleeping period of the site and it is carried out if and only if the replica is dominant by its number of version compared to the target. In the contrary case the propagation is refused. In practice, the replica source diffuses its updates with the other replicas of the same site;

iv) Detection and resolution of the conflicts intra-site: If two versions of two metadata different are identical then a conflict is detected between two replicas.

4.2.2 What is the global consistency?

Global consistency is also called consistency inter-sites. Its principal objective is to ensure consistency between the various replicas of the same data of the Grid, which corresponds to make converge the replicas towards a reference replica for a Grid and it is founded on the pessimistic approach of replication. This reference replica will transmit its information towards the other representatives of the nodes. The Moment of release of global consistency Global consistency can be launched according to several situations:

1. If the account of conflicts of a site exceeds a certain threshold, then a rate of inconsistencies is very high, in this case the site becomes unable to correctly serve the requests of the clients, we will speak about a divergence of the replicas according to a local view;
2. If the average of account of conflicts of the whole of sites exceeds a certain threshold, that corresponds to a divergence of the copies according to a global view;
3. If the distance between two copies of intra-site or inter-sites reaches a breaking value, which corresponds to the margin between two replicas of the same data;
4. If the rate of writing reaches a given value;
5. After each past period (periodically).

5 Measuring our approach

A metric is a quantity related to the performance and availability of the Grid. There are several measures which can be considered in the evaluation of Grid consistency strategies [3, 19]. The quality of service of optimistic replication systems in the face of updates is the degree to which the system presents an illusion of connectivity to a single up-to-date copy of all objects to all users. In real replicated systems, this illusion must necessarily be violated, and quantifying the user-visible effects of such violations is key to determining how well the system performs. In this

work, we consider several types of measures to the evaluation of our approach with the approaches pessimistic and optimistic. In general, these measurements can represent how fast or reliability a service is served. The first type represents the quality of rendered service, the second type of measurement is used to estimate the response time.

5.1 Metrics of Quality of Service

In this category of measurements, we can define several measurements of quality of service, for example:

- Count Conflicts: this metric can represent by several manners:
  - Count of conflicts by time unit: its principle is to count conflicts number of all sites in the grid at each time period;
  - Conflicts count by sites: the principal goal by the use of this metric is to show the master choice and the number of nodes in conflicts counts;
  - Conflicts count by sites number: it is a conflict count for different number of sites, this metric makes it possible to study the behavior of the number of conflicts when the number of sites increases in a continuous way;
- Distance from replicas: the distance is related to the number of update of replicas, it is equal to the difference between the maximum number of update and the minimum of an object.

  Four metrics can be derived from the distance metric:
  - It consists to count the distance of all sites in the grid, at each time period. This metric makes it possible to give a periodic sight of the margin of the replicas to each selected moment;
  - Distance by sites: if we have a sites group, the distance by sites consists to count the distance for each site at the end of simulation. This measurement makes it possible to have a total sight of the margin for the sites unit at the end of simulation;
  - Mean distance by sites: Simply it is the distance of every site by sites number. This metric makes it possible to inform us on the average state from point of view outdistances average inside the grid;
  - Mean distance by sites number: It is the distance by sites number for different number of sites. This measurement seems to us very interesting to study the margins of the replicas compared to the evolution of a number of the sites of the grid.

5.2 Metrics of Performance and Availability

This category of metric is very related to the physical characteristics of the grid, such as, the band-width, the speed of the elements of calculations, the capacity of the elements of storage, etc In general, we define two types of measurements for this category, often with very high Coefficient of correlation:

1. Performance: generally the performance with an indicator of performance: response time of requests, quantity of requests to be treated per unit of time, quantity of resources used, etc. Most of the time, we consider that a proposal contributing to the improvement of the one of these indicators, generates a reduction of the costs and implicitly an improvement of the performances;
2. Availability: allows to ensure accessibility the resources as that is necessary. Often, this measurement is associated the latencies of the requests of the clients, cuts queue of a site or a computing element, etc.

6 Simulation results

As it was already discussed, the hybrid approach suggested is implemented in the environment of simulation of the grids of data OptorSim. For our simulation, several parameters can be taken, according in the following table (*Table. 1*).

Notation	Definition
k	Number of Sites
$Site_j$	Number of (CE,SE)
NQ	Total number of requests
Ta_i	Arrival time of $request_i$
F_{ji}	Reliability $(CE, SE)_i$ of $site_j$
CR	Size of replica
Bd_j	Network Bandwidth of $site_j$

Table 1. Parameters for simulation

Figure *4* shows that the hybrid approach contains a number of conflicts much lower than the optimistic approach.

Conflicts count by sites number is a conflict count for different number of sites, we observe in the figure *5* that when we increase the sites number, conflicts number increase for the two approaches (*Optimistic, Hybrid*) until the end of simulation, but the result of the hybrid approach proposed are better than the optimistic approach.

Figure *6* shows merely that this distance is very significant in the optimistic approach compared to our approach.

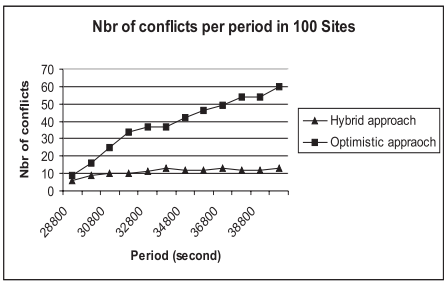

Figure 4. Count conflicts by period

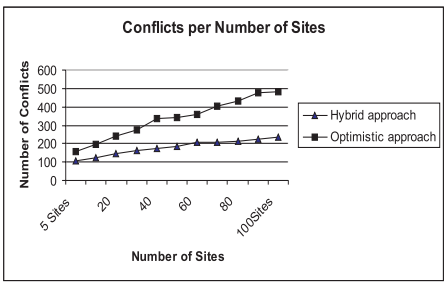

Figure 5. Conflicts count by sites number

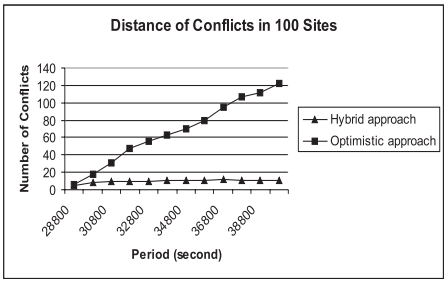

Figure 6. Distance by Period

We observe in figures *4, 5 and 5*, that the divergence is very fast in the optimistic approach one.

In order to study the performance and the availability, we chose to compare our approach with the two protocols of pessimistic consistency: *ROWA* (Read One Write All) and *majority Quorum* [2, 12]. The results of simulation shown in *Figures 7 and 8* prove that the protocol suggested gives better results compared to the two pessimistic protocols. We notice that protocol *ROWA* very quickly becomes impracticable when the number of sites increases, the consequence is that it is an unsuitable protocol to large scale systems.

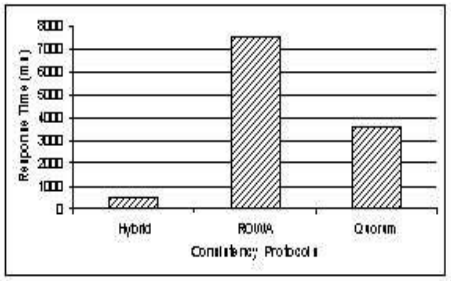

Figure 7. Average response time of 20 Sites

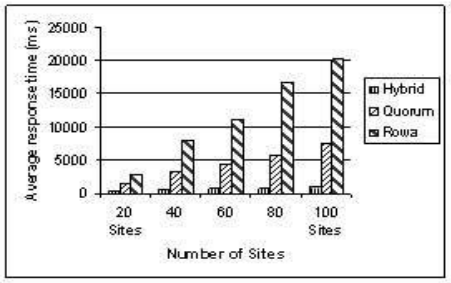

Figure 8. Average response time per number of sites

The figure *9* summarizes the percentages of the profits to be gained by the hybrid approach compared to the two techniques ROWA and Quorum. We notice, that the profit to be gained can go up to 91% to the profit of ROWA and 77% to the profit of the Quorum approach.

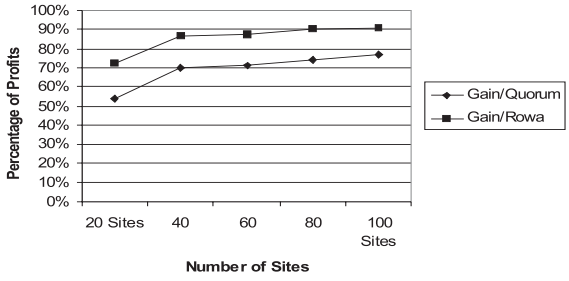

Figure 9. Comparison with Hybrid and Pessimistic approaches

7 Conclusions and future works

In this paper we described the hybrid approach for consistency management in large scale system. Our hybrid approach is based on a model on two levels implemented in the OptorSim environment, whose objective is double vacation, because it initially makes it possible to reduce response times compared to a completely Pessimistic approach and in the second time to improve the quality of service compared to a totally Optimistic. The results of simulation obtained are very satisfactory, and that the approach is very promising for the application requiring certain quality of service and acceptable response time. The results of simulation obtained are very satisfactory, and that the suggested approach is very promising for the large scale application requiring a certain quality of service.
The results of simulation obtained are very satisfactory, and show that the approach is very promising especially when requiring quality of service and acceptable response time. Some works can be led to the future as:

- Experimentation of the proposed approach on a real grid;
- Supply the layer1 a multi-agents system to decide on the choice of the Global reference replica;

References

[1] Edgsim: A simulation of the european datagrid.

[2] Y. Amir and A. Wool. Optimal availability quorum systems: Theory and practice. *Information Processing Letters*, 65(5):223–228, 1998.

[3] G. Belalem and Y. Slimani. A hybrid approach for consistency management in large scale systems. In I. C. Society, editor, *International Conference on Networking and Services (ICNS'06)*, volume 0, pages 71–71, Silicon Valley, USA, 16-19 July 2006.

[4] W. Bell, D. Cameron, R. Carvajal-Schiaffino, P. Millar, C.Nicholson, K. Stockinger, and F. Zini. *OptorSim v1.0 Installation and User Guide*, February 2004.

[5] W. H. Bell, G. D. Cameron, L. Capozza, A. P. Millar, K. Stockinger, and F. Zini. Optorsim : A grid simulator for studying dynamic data replication strategies. *Int. Journal of High Performance Computing Applications*, 17(4):403–416, 2003.

[6] R. Buyya and M. Murshed. Gridsim: A toolkit for the modeling and simulation of distributed resource management and scheduling for grid computing. *Journal of Concurrency and Computation: Practice and Experience (CCPE)*, 14(13–15):1175–1220, 2002.

[7] D. G. Cameron, A. P. Millar, C. Nicholson, R. Carvajal-Schiaffino, K. Stockinger, and F. Zini. Analysis of scheduling and replica optimisation strategies for data grids using optorsim. *Journal of Grid Computing*, 2(1):57–69, 2004.

[8] H. Casanova, A. Legrand, and L. Marchal. Scheduling distributed applications: the simgrid simulation framework. In *In Proceedings of the Third IEEE International Symposium on Cluster Computing and the Grid (CCGrid'03)*, pages 138–144, Tokyo, Japan, May 2003.

[9] A. Domenici, F. Donno, G. Pucciani, H. Stockinger, and K. Stockinger. Replica consistency in a data grid. *Nuclear Instruments and Methods in Physics Research A*, 534:24–28, 2004.

[10] C. Dumitrescu and I. Foster. Gangsim: A simulator for grid scheduling studies. In *in Proceedings of the IEEE International Symposium on Cluster Computing and the Grid (CCGrid'05)*, pages 1151–1158, Cardiff, UK, May 2005.

[11] I. Foster and C. Kesselman. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kauffman Publishers Inc., San Francisco, 1999.

[12] S. Goel, H. Sharda, and D. Taniar. Replica synchronisation in grid databases. *Int. J. Web and Grid Services*, 1(1):87–112, 2005.

[13] J. Gray, P. Helland, P. O. Neil, and D. Shasha. The dangers of replication and a solution. In *ACM SIGMOD International Conference on Management of Data*, pages 173–182, Montreal, Quebec, Canada, 4-5 June 1996. ACM Press.

[14] K. Ranganathan and I. Foster. Identifying dynamic replication strategies for a high-performance data grid. In S. Berlin, editor, *Grid: Second International Workshop*, volume 2242, pages 75–86, Denver, CO, USA, 12 November 2001.

[15] K. Ranganathan and I. Foster. Decoupling computation and data scheduling in distributed data-intensive applications. In *International Symposium of High Performance Distributed Computing*, pages 352–358, Edinburgh, Scotland, UK, 2002.

[16] Y. Saito and M. Shapiro. Optimistic replication. *ACM Comput. Surv.*, 37(1):42–81, 2005.

[17] H. Song, X. Liu, and D. Jakobsen. The microgrid: ascientific tool for modeling computational grids. In *IEEE Supercomputing*, November 2000.

[18] A. Takefusa, S. Matsuoka, K. Aida, H. Nakada, and U. Nagashima. Overview of a performance evaluation system for global computing scheduling algorithms. In *HPDC'99: Proceedings of the Eighth IEEE International Symposium on High Performance Distributed Computing*, pages 11–17, Washington, DC, USA, 1999.

[19] A.-I. Wang, P. L.Reiher, R. Bagrodia, and G. H. Kuenning. Understanding the behavior of the conflict-rate metric in optimistic peer replication. In I. C. Society, editor, *DEXA '02: Proceedings of the 13th International Workshop on Database and Expert Systems Applications*, pages 757–764, Washington, DC, USA, 2002.