Eduardo Tovar
Philippas Tsigas
Hacène Fouchal (Eds.)

LNCS 4878

Principles of Distributed Systems

11th International Conference, OPODIS 2007
Guadeloupe, French West Indies, December 2007
Proceedings

Springer

Eduardo Tovar Philippas Tsigas
Hacène Fouchal (Eds.)

Principles of Distributed Systems

11th International Conference, OPODIS 2007
Guadeloupe, French West Indies, December 17-20, 2007
Proceedings

Springer

Volume Editors

Eduardo Tovar
Polytechnic Institute of Porto (ISEP-IPP)
Department of Computer Engineering
Rua Dr. António Bernardino de Almeida 431, 4200-072 Porto, Portugal
E-mail: emt@dei.isep.ipp.pt

Philippas Tsigas
Chalmers University of Technology
Department of Computer Science & Engineering
412 96 Göteborg, Sweden
E-mail: tsigas@chalmers.se

Hacène Fouchal
GRIMAAG, Université des Antilles et de Guyane
97157 Pointe-à-Pitre, Guadeloupe, France
E-mail: Hacene.Fouchal@univ-ag.fr

Library of Congress Control Number: 2007940869

CR Subject Classification (1998): C.2.4, D.1.3, D.2.7, D.2.12, D.4.7, C.3

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

ISSN	0302-9743
ISBN-10	3-540-77095-X Springer Berlin Heidelberg New York
ISBN-13	978-3-540-77095-4 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12199679 06/3180 5 4 3 2 1 0

Lecture Notes in Computer Science 4878

Preface

It is our pleasure to welcome you to the 11th International Conference on Principles of Distributed Systems (OPODIS 2007), held during December 17–20, 2007, in Guadeloupe, French West Indies.

During the past years, OPODIS has established itself as one of the most important events related to principles of distributed computing, networks and systems. In this year's edition, we received 106 submissions in response to the call for papers. Papers were sought soliciting original research contributions to the theory, specifications, design and implementation of distributed systems, including: communication and synchronization protocols; distributed algorithms, multiprocessor algorithms; distributed cooperative computing; embedded systems; fault-tolerance, reliability, availability; grid and cluster computing; location- and context-aware systems; mobile agents and autonomous robot; mobile computing and networks; peer-to-peer systems, overlay networks; complexity and lower bounds; performance analysis of distributed systems; real-time systems; security issues in distributed computing and systems; sensor networks: theory and practice; specification and verification of distributed systems; testing and experimentation with distributed systems.

It was a hard task to select the 32 excellent papers that are compiled in this volume. All submissions received at least three reviews, resulting in an overall number of more than 350 reviews—involving more than 100 reviewers—being conducted. An important piece of the process was the Program Committee electronic meeting held during the week of September 10. It was definitely an outstanding performance by all 42 Program Committee members and their co-reviewers. We are convinced that a very good set of papers was selected for presentation at OPODIS 2007.

This year's edition also featured two exciting keynote talks by Moti Yung (Google Inc., USA) and Tarek Abdelzaher (University of Illinois at Urbana-Champaign, USA). We are grateful that those two distinguished speakers accepted our invitation to share with us their views.

It would not have been possible to set up this exciting program without the close cooperation and support of the General Chair, Hacène Fouchal, who deserves a big share of credit. We would also like to thank Thibault Bernard for his role as Publicity Chair and on managing both the electronic submission and reviewing system and OPODIS 2007 Web site, and to the other members of the Organizing Committee, Céline Butelle, Harry Gros-Désormeaux and Vincent Levorato.

December 2007

Eduardo Tovar
Philippas Tsigas

Conference Organization

OPODIS 2007 was organized by "Université des Antilles et de la Guyane" in Guadeloupe, French West Indies and APODIS "Association Pour la Diffusion Scientifique."

General Chair

Hacène Fouchal	Université des Antilles et de la Guyane, France

Program Co-chairs

Eduardo Tovar	Polytechnic Institute of Porto, Portugal
Philippas Tsigas	Chalmers University of Technology, Sweden

Program Committee

Tarek Abdelzaher	University of Illinois at Urbana-Champaign, USA
Mustaque Ahamad	Georgia Institute of Technology, USA
James Anderson	University of North Carolina at Chapel Hill, USA
Bjorn Andersson	Polytechnic Institute of Porto, Portugal
Alain Bui	University of Reims, France
Leandro Buss Becker	University of Santa Catarina, Brazil
Franck Cappello	LRI, France
Ajoy Datta	University of Nevada at Las Vegas, USA
Murat Demirbas	State University of New York at Buffalo, USA
Shlomy Dolev	Ben-Gurion University, Israel
Pascal Felber	University of Neuchatel, Switzerland
Gerhard Fohler	University of Kaiserslautern, Germany
Marisol Garcia Valls	UC3 Madrid, Spain
Christopher Gill	Washington University at St. Louis, USA
Rachid Guerraoui	EPFL, Switzerland
Maurice Herlihy	Brown University, USA
Ted Herman	University of Iowa, USA
Philippe Hunel	Université des Antilles et de la Guyane, France
Pierre Jansen	University of Twente, The Netherlands
Vana Kalogeraki	University of California at Riverside, USA
Anne-Marie Kermarrec	INRIA, France

Herman Kopetz	Technical University of Vienna, Austria
Steve Liu	Texas A&M University, USA
Lucia Lo Bello	University of Catania, Italy
Maged Michael	IBM, USA
Rami Melhem	University of Pittsburgh, USA
Yves Metivier	LaBRI, France
Christian Poellabauer	University of Notre Dame, USA
Raj Rajkumar	Carnegie Mellon University, USA
Krithi Ramamritham	Indian Institute of Technology Bombay, India
Binoy Ravindran	Virginia Tech, USA
Luis Rodrigues	University of Lisbon, Portugal
Nicola Santoro	Carleton University, Canada
Christian Scheideler	Technical University of Munich, Germany
Marc Shapiro	INRIA, France
Paul Spirakis	CTI, Greece
Alex Shvartsman	University of Connecticut, USA
Sebastien Tixeuil	LRI, France
Carlos Becker Westphall	University of Santa Catarina, Brazil
Peter Widmayer	ETH Zurich, Switzerland
Masafumi Yamashita	Kyushu University, Japan
Nina Yevtushenko	Tomsk State University, Russia

Organizing Committee

Thibault Bernard	Université de Reims Champagne-Ardenne, France
Céline Butelle	Université de Paris 8, France
Harry Gros-Désormeaux	Université des Antilles et de la Guyane, France
Vincent Levorato	EPHE, Sorbonne, France

Publicity Chair

Thibault Bernard	Université de Reims Champagne-Ardenne, France

Steering Committee

Alain Bui	Université de Reims Champagne-Ardenne, France
Marc Bui	EPHE, Sorbonne, France
Hacène Fouchal	Université des Antilles et de la Guyane, France
Roberto Gomez	ITESM-CEM, Mexico
Nicola Santoro	Carleton University, Canada
Philippas Tsigas	Chalmers University of Technology, Sweden

Reviewers

Tarek Abdelzaher
Mustaque Ahamad
Murat Ali Bayir
James Anderson
Bjorn Andersson
Carlos Becker Westphall
Doina Bein
Thibault Bernard
Marin Bertier
Olga Brukman
Alain Bui
Leandro Buss Becker
John Calandrino
Franck Cappello
David Cash
Ioannis Chatzigiannakis
Tim Childress
Ajoy Datta
Seda Davtyan
Murat Demirbas
UmaMaheswari Devi
Stphane Devismes
Martin Dietzfelbinger
Shlomi Dolev
Ji (Adam) Dou
Yannis Drougas
Mahmoud Elhaddad
Sherif Fahmy
Gilles Fedak
Pascal Felber
Hen Fitoussi
Gerhard Fohler
Marisol Garcia Valls
Chris Gill
Sameh Gobrial
Maria Gradinariu
Harry Gros-Desormeaux
Maxim Gromov
Rachid Guerraoui
Arobinda Gupta
Kai Han
Nicolas Hanusse
Thomas Hérault
Maurice Herlihy
Ted Herman
Philippe Hunel
Muzammil Hussain
Mathieu Jan
Pierre Jansen
Rajeev Joshi
Vana Kalogeraki
Ronen Kat
Sotirios Kentros
Anne-Marie Kermarrec
Sherif Khattab
Ralf Klasing
Derrick Kondo
Herman Kopetz
Mimor Lahiani
Lawrence Larmore
Joao Leitao
Hennadiy Leontyev
Vasia Liagkou
Steve Liu
Lucia Lo Bello
Xuming Lu
Mihai Marin-Perianu
Charles Martel
Pau Marti
Rami Melhem
Yves Metivier
Maged Michael
Hugo Miranda
Jose Mocito
Peter Musial
Gil Neiger
Nicolas Nicolaou
Florent Nolot
Lucia Penso
Sylvain Peyronnet
Christian Poellabauer
Giuseppe Prencipe
Svetlana Prokopenko
Cyril Rabat
Raj Rajkumar
Krithi Ramamritham
Binoy Ravindran
Torvald Riegel
Luis Rodrigues
Rodrigo Rodrigues
Nasser Saheb
Nicola Santoro
Christian Scheideler
Andrey Shabaldin
Natalya Shabaldina
Marc Shapiro
Alex Shvartsman
Marcelo Sobral
Devan Sohier
Elaine Sonderegger
Onur Soysal
Paul Spirakis
Luiz-Angelo Steffenel
Pierre Sutra
Hisao Tamaki
Sebastien Tixeuil
Emanuele Toscano
Nir Tzachar
Nicolas Vidot
Zhijum Wang
Peter Widmayer
Ruibin Xu
Reuven Yagel
Masafumi Yamashita
Nina Yevtushenko

Sponsoring Institutions

UAG, Université des Antilles et de la Guyane
EPHE, l'Ecole Pratique des Hautes Etudes, Sorbonne, France
LaISC, Laboratoire d'Informatique et des Systèmes Complexes, France
GDR ARP du CNRS
INRIA, Institut National de la Recherche en Informatique et en Automatique

Lecture Notes in Computer Science

Sublibrary 1: Theoretical Computer Science and General Issues

For information about Vols. 1–4510
please contact your bookseller or Springer

Vol. 4878: E. Tovar, P. Tsigas, H. Fouchal (Eds.), Principles of Distributed Systems. XIII, 457 pages. 2007.

Vol. 4873: S. Aluru, M. Parashar, R. Badrinath, V.K. Prasanna (Eds.), High Performance Computing – HiPC 2007. XXIV, 663 pages. 2007.

Vol. 4863: A. Bonato, F.R.K. Chung (Eds.), Algorithms and Models for the Web-Graph. X, 217 pages. 2007.

Vol. 4855: V. Arvind, S. Prasad (Eds.), FSTTCS 2007: Foundations of Software Technology and Theoretical Computer Science. XIV, 558 pages. 2007.

Vol. 4847: M. Xu, Y. Zhan, J. Cao, Y. Liu (Eds.), Advanced Parallel Processing Technologies. XIX, 767 pages. 2007.

Vol. 4846: I. Cervesato (Ed.), Advances in Computer Science – ASIAN 2007. XI, 313 pages. 2007.

Vol. 4838: T. Masuzawa, S. Tixeuil (Eds.), Stabilization, Safety, and Security of Distributed Systems. XIII, 409 pages. 2007.

Vol. 4835: T. Tokuyama (Ed.), Algorithms and Computation. XVII, 929 pages. 2007.

Vol. 4783: J. Holub, J. Žďárek (Eds.), Implementation and Application of Automata. XIII, 324 pages. 2007.

Vol. 4782: R. Perrott, B.M. Chapman, J. Subhlok, R.F. de Mello, L.T. Yang (Eds.), High Performance Computing and Communications. XIX, 823 pages. 2007.

Vol. 4771: T. Bartz-Beielstein, M.J. Blesa Aguilera, C. Blum, B. Naujoks, A. Roli, G. Rudolph, M. Sampels (Eds.), Hybrid Metaheuristics. X, 202 pages. 2007.

Vol. 4770: V.G. Ganzha, E.W. Mayr, E.V. Vorozhtsov (Eds.), Computer Algebra in Scientific Computing. XIII, 460 pages. 2007.

Vol. 4763: J.-F. Raskin, P.S. Thiagarajan (Eds.), Formal Modeling and Analysis of Timed Systems. X, 369 pages. 2007.

Vol. 4746: A. Bondavalli, F. Brasileiro, S. Rajsbaum (Eds.), Dependable Computing. XV, 239 pages. 2007.

Vol. 4743: P. Thulasiraman, X. He, T.L. Xu, M.K. Denko, R.K. Thulasiram, L.T. Yang (Eds.), Frontiers of High Performance Computing and Networking ISPA 2007 Workshops. XXIX, 536 pages. 2007.

Vol. 4742: I. Stojmenovic, R.K. Thulasiram, L.T. Yang, W. Jia, M. Guo, R.F. de Mello (Eds.), Parallel and Distributed Processing and Applications. XX, 995 pages. 2007.

Vol. 4739: R. Moreno Díaz, F. Pichler, A. Quesada Arencibia (Eds.), Computer Aided Systems Theory – EUROCAST 2007. XIX, 1233 pages. 2007.

Vol. 4736: S. Winter, M. Duckham, L. Kulik, B. Kuipers (Eds.), Spatial Information Theory. XV, 455 pages. 2007.

Vol. 4732: K. Schneider, J. Brandt (Eds.), Theorem Proving in Higher Order Logics. IX, 401 pages. 2007.

Vol. 4731: A. Pelc (Ed.), Distributed Computing. XVI, 510 pages. 2007.

Vol. 4726: N. Ziviani, R. Baeza-Yates (Eds.), String Processing and Information Retrieval. XII, 311 pages. 2007.

Vol. 4711: C.B. Jones, Z. Liu, J. Woodcock (Eds.), Theoretical Aspects of Computing – ICTAC 2007. XI, 483 pages. 2007.

Vol. 4710: C.W. George, Z. Liu, J. Woodcock (Eds.), Domain Modeling and the Duration Calculus. XI, 237 pages. 2007.

Vol. 4708: L. Kučera, A. Kučera (Eds.), Mathematical Foundations of Computer Science 2007. XVIII, 764 pages. 2007.

Vol. 4707: O. Gervasi, M.L. Gavrilova (Eds.), Computational Science and Its Applications – ICCSA 2007, Part III. XXIV, 1205 pages. 2007.

Vol. 4706: O. Gervasi, M.L. Gavrilova (Eds.), Computational Science and Its Applications – ICCSA 2007, Part II. XXIII, 1129 pages. 2007.

Vol. 4705: O. Gervasi, M.L. Gavrilova (Eds.), Computational Science and Its Applications – ICCSA 2007, Part I. XLIV, 1169 pages. 2007.

Vol. 4703: L. Caires, V.T. Vasconcelos (Eds.), CONCUR 2007 – Concurrency Theory. XIII, 507 pages. 2007.

Vol. 4700: C.B. Jones, Z. Liu, J. Woodcock (Eds.), Formal Methods and Hybrid Real-Time Systems. XVI, 539 pages. 2007.

Vol. 4699: B. Kågström, E. Elmroth, J. Dongarra, J. Waśniewski (Eds.), Applied Parallel Computing. XXIX, 1192 pages. 2007.

Vol. 4698: L. Arge, M. Hoffmann, E. Welzl (Eds.), Algorithms – ESA 2007. XV, 769 pages. 2007.

Vol. 4697: L. Choi, Y. Paek, S. Cho (Eds.), Advances in Computer Systems Architecture. XIII, 400 pages. 2007.

Vol. 4688: K. Li, M. Fei, G.W. Irwin, S. Ma (Eds.), Bio-Inspired Computational Intelligence and Applications. XIX, 805 pages. 2007.

Vol. 4684: L. Kang, Y. Liu, S. Zeng (Eds.), Evolvable Systems: From Biology to Hardware. XIV, 446 pages. 2007.

Vol. 4683: L. Kang, Y. Liu, S. Zeng (Eds.), Advances in Computation and Intelligence. XVII, 663 pages. 2007.

Vol. 4681: D.-S. Huang, L. Heutte, M. Loog (Eds.), Advanced Intelligent Computing Theories and Applications. XXVI, 1379 pages. 2007.

Vol. 4672: K. Li, C. Jesshope, H. Jin, J.-L. Gaudiot (Eds.), Network and Parallel Computing. XVIII, 558 pages. 2007.

Vol. 4671: V.E. Malyshkin (Ed.), Parallel Computing Technologies. XIV, 635 pages. 2007.

Vol. 4669: J.M. de Sá, L.A. Alexandre, W. Duch, D. Mandic (Eds.), Artificial Neural Networks – ICANN 2007, Part II. XXXI, 990 pages. 2007.

Vol. 4668: J.M. de Sá, L.A. Alexandre, W. Duch, D. Mandic (Eds.), Artificial Neural Networks – ICANN 2007, Part I. XXXI, 978 pages. 2007.

Vol. 4666: M.E. Davies, C.J. James, S.A. Abdallah, M.D. Plumbley (Eds.), Independent Component Analysis and Blind Signal Separation. XIX, 847 pages. 2007.

Vol. 4665: J. Hromkovič, R. Královič, M. Nunkesser, P. Widmayer (Eds.), Stochastic Algorithms: Foundations and Applications. X, 167 pages. 2007.

Vol. 4664: J. Durand-Lose, M. Margenstern (Eds.), Machines, Computations, and Universality. X, 325 pages. 2007.

Vol. 4661: U. Montanari, D. Sannella, R. Bruni (Eds.), Trustworthy Global Computing. X, 339 pages. 2007.

Vol. 4649: V. Diekert, M.V. Volkov, A. Voronkov (Eds.), Computer Science – Theory and Applications. XIII, 420 pages. 2007.

Vol. 4647: R. Martin, M.A. Sabin, J.R. Winkler (Eds.), Mathematics of Surfaces XII. IX, 509 pages. 2007.

Vol. 4646: J. Duparc, T.A. Henzinger (Eds.), Computer Science Logic. XIV, 600 pages. 2007.

Vol. 4644: N. Azémard, L. Svensson (Eds.), Integrated Circuit and System Design. XIV, 583 pages. 2007.

Vol. 4641: A.-M. Kermarrec, L. Bougé, T. Priol (Eds.), Euro-Par 2007 Parallel Processing. XXVII, 974 pages. 2007.

Vol. 4639: E. Csuhaj-Varjú, Z. Ésik (Eds.), Fundamentals of Computation Theory. XIV, 508 pages. 2007.

Vol. 4638: T. Stützle, M. Birattari, H. H. Hoos (Eds.), Engineering Stochastic Local Search Algorithms. X, 223 pages. 2007.

Vol. 4630: H.J. van den Herik, P. Ciancarini, H.H.L.M.(J.) Donkers (Eds.), Computers and Games. XII, 283 pages. 2007.

Vol. 4628: L.N. de Castro, F.J. Von Zuben, H. Knidel (Eds.), Artificial Immune Systems. XII, 438 pages. 2007.

Vol. 4627: M. Charikar, K. Jansen, O. Reingold, J.D.P. Rolim (Eds.), Approximation, Randomization, and Combinatorial Optimization. XII, 626 pages. 2007.

Vol. 4624: T. Mossakowski, U. Montanari, M. Haveraaen (Eds.), Algebra and Coalgebra in Computer Science. XI, 463 pages. 2007.

Vol. 4623: M. Collard (Ed.), Ontologies-Based Databases and Information Systems. X, 153 pages. 2007.

Vol. 4621: D. Wagner, R. Wattenhofer (Eds.), Algorithms for Sensor and Ad Hoc Networks. XIII, 415 pages. 2007.

Vol. 4619: F. Dehne, J.-R. Sack, N. Zeh (Eds.), Algorithms and Data Structures. XVI, 662 pages. 2007.

Vol. 4618: S.G. Akl, C.S. Calude, M.J. Dinneen, G. Rozenberg, H.T. Wareham (Eds.), Unconventional Computation. X, 243 pages. 2007.

Vol. 4616: A.W.M. Dress, Y. Xu, B. Zhu (Eds.), Combinatorial Optimization and Applications. XI, 390 pages. 2007.

Vol. 4614: B. Chen, M. Paterson, G. Zhang (Eds.), Combinatorics, Algorithms, Probabilistic and Experimental Methodologies. XII, 530 pages. 2007.

Vol. 4613: F.P. Preparata, Q. Fang (Eds.), Frontiers in Algorithmics. XI, 348 pages. 2007.

Vol. 4600: H. Comon-Lundh, C. Kirchner, H. Kirchner (Eds.), Rewriting, Computation and Proof. XVI, 273 pages. 2007.

Vol. 4599: S. Vassiliadis, M. Bereković, T.D. Hämäläinen (Eds.), Embedded Computer Systems: Architectures, Modeling, and Simulation. XVIII, 466 pages. 2007.

Vol. 4598: G. Lin (Ed.), Computing and Combinatorics. XII, 570 pages. 2007.

Vol. 4596: L. Arge, C. Cachin, T. Jurdziński, A. Tarlecki (Eds.), Automata, Languages and Programming. XVII, 953 pages. 2007.

Vol. 4595: D. Bošnački, S. Edelkamp (Eds.), Model Checking Software. X, 285 pages. 2007.

Vol. 4590: W. Damm, H. Hermanns (Eds.), Computer Aided Verification. XV, 562 pages. 2007.

Vol. 4588: T. Harju, J. Karhumäki, A. Lepistö (Eds.), Developments in Language Theory. XI, 423 pages. 2007.

Vol. 4583: S.R. Della Rocca (Ed.), Typed Lambda Calculi and Applications. X, 397 pages. 2007.

Vol. 4580: B. Ma, K. Zhang (Eds.), Combinatorial Pattern Matching. XII, 366 pages. 2007.

Vol. 4576: D. Leivant, R. de Queiroz (Eds.), Logic, Language, Information and Computation. X, 363 pages. 2007.

Vol. 4547: C. Carlet, B. Sunar (Eds.), Arithmetic of Finite Fields. XI, 355 pages. 2007.

Vol. 4546: J. Kleijn, A. Yakovlev (Eds.), Petri Nets and Other Models of Concurrency – ICATPN 2007. XI, 515 pages. 2007.

Vol. 4545: H. Anai, K. Horimoto, T. Kutsia (Eds.), Algebraic Biology. XIII, 379 pages. 2007.

Vol. 4533: F. Baader (Ed.), Term Rewriting and Applications. XII, 419 pages. 2007.

Vol. 4528: J. Mira, J.R. Álvarez (Eds.), Nature Inspired Problem-Solving Methods in Knowledge Engineering, Part II. XXII, 650 pages. 2007.

Vol. 4527: J. Mira, J.R. Álvarez (Eds.), Bio-inspired Modeling of Cognitive Tasks, Part I. XXII, 630 pages. 2007.

Vol. 4525: C. Demetrescu (Ed.), Experimental Algorithms. XIII, 448 pages. 2007.

Vol. 4514: S.N. Artemov, A. Nerode (Eds.), Logical Foundations of Computer Science. XI, 513 pages. 2007.

Vol. 4513: M. Fischetti, D.P. Williamson (Eds.), Integer Programming and Combinatorial Optimization. IX, 500 pages. 2007.

Table of Contents

A Decentralized, Scalable, and Autonomous Grid Monitoring System

Laurent Baduel[1] and Satoshi Matsuoka[1,2]

[1] Tokyo Institute of Technology,
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
[2] National Institut of Informatics,
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan
baduel@smg.is.titech.ac.jp, matsu@is.titech.ac.jp

Abstract. Grid monitoring systems collect a substantial amount of information on the infrastructure's status in order to perform various tasks, more commonly to provide a better use of the grid's entities. Modern computational and data grids have become very complex by their size, their heterogeneity, their interconnection. Monitoring systems as any other grid's tools have to adapt to this evolution. In this paper we present a decentralized, scalable, and autonomous grid monitoring system able to tackle the growths of scale and complexity. System's components communications are hierarchically organized on a peer-to-peer overlay network. Fresh information is efficiently propagated thanks to an *directed* gossip protocol that limits the number of message. Automation of key management operations eases system administration and maintenance. This approach provides scalability and adaptability. The main properties of our application are presented and discussed. Performance measurements confirm the efficiency of our system.

1 Introduction

Grid platforms with their ever-growing communication infrastructures and computing applications become larger and larger, which results in an exponential complexity in their engineering and maintenance operations. To efficiently handle such large and complex systems monitoring is necessary. By providing a global view of the system monitoring tools allow identifying performance problems and assisting in resources scheduling. Modern large scale systems do not allow anymore centralized organizations, with hand deployment, configuration, and administration. Automation of key operations must be introduced in such systems to free the administrators and programmers of many tiresome tasks.

After presenting related work, we propose a grid monitoring system built to address these challenges. It provides a scalable and portable monitoring of a wide range of entities connected in distributed systems. This decentralized tool achieves its communications thanks to a peer-to-peer overlay network. Peer-to-peer has become a popular way to communicate on grid thanks to its scalability, its decentralization, and its resistance to faults. It has already proved their efficiency in many aspects of distributed applications such as embarrassingly parallel computing, persistent and scalable storage, and especially in file sharing and dissemination. We use a gossip protocol to quickly spread

E. Tovar, P. Tsigas, and H. Fouchal (Eds.): OPODIS 2007, LNCS 4878, pp. 1–15, 2007.

information in the entire system. Components of the system are organized as a directed acyclic graph to guide information from their source to the bases storing the entire system state. This, combined to over-aged information filtering, reduces the amount of exchanged messages. Finally we have deployed a first implementation on a real grid and evaluated performance of our system.

In summary the contributions of this article are *(1)* the details of the conception of a decentralized and scalable grid monitoring system in which components are hierarchically organized through a directed acyclic graph on a peer-to-peer overlay network that provides a valuable communication layer thanks to a gossip multicast protocol, assures good performance on grids, and allows self-management of the system; *(2)* a performance evaluation of this system driven in a real large-sized grid. Speed of information dissemination, age of recorded information, impact of messages limitation, and dynamic adaptability are examined and discussed. This paper rather focuses on the fast and decentralized dissemination of information than on other aspects of monitoring such as sensors implementation or database organization.

The rest of this article is organized as follow: Section 2 describes the general properties of a monitoring system and insists on the specific requirements for grid environments. Then it presents related work. Section 3 introduces the architecture for monitoring system on which we based our implementation. Section 4 presents our original communication scheme based on peer-to-peer and what we name a *directed* gossip protocol. Then Section 5 details self-management mechanisms introduced in the system to help scalability and maintenance. Section 6 details the implementation and presents performance evaluations of the application. Finally Section 7 concludes the article and presents expected future works.

2 Grid Monitoring Systems

The activity of measuring significant resources parameters allows analyzing the usage, the behavior, and the performance of a cluster or a grid. It also provides Grid monitoring systems to collect a substantial amount of information on the infrastructure's status in order to perform various tasks, more commonly to provide a better use of the grid's entities.

Monitoring a system consists of observing events and communicating them to who are interested in that information. There are commonly two systems on a grid. According to [1], a *grid monitoring system* manages rapidly changing status information, such as the load of a CPU or the throughput of a network link. The high dynamicity of data that a grid monitoring system must handle makes it different from a *grid information system* which handles more static data, for instance the hardware configuration of a node. Although grid infrastructures can benefit a lot from a unified system that handles both roles. The global knowledge provided by a unified grid monitoring system helps resource scheduling, allocation and usage: reservation tools can be plugged in order to distribute resources and guarantee quality of service.

Monitoring dedicated to grids is subject to a growing interest. The recent large grids do not support anymore efficiently the existing monitoring tools. Most of existing solutions are adaptations of cluster oriented monitoring tools, and then lay to scalability and robustness issues. As detailed further the Network Weather Service is built around a centralized controller, and the Globus Monitoring and Discovery System suffers performance and scalability issues due to its LDAP architecture. On the contrary, the grid monitoring system we propose is adapted to the grid thanks to its scalability and fault tolerance ability, and also thank to its autonomous management.

The rest of this section presents a survey about grid monitoring systems and their potential autonomous mechanisms for configuration and adaptation.

2.1 Network Weather Service (NWS)

The Network Weather Service [2] is a distributed system that periodically monitors and dynamically forecasts the performance that various network and computational resources can deliver over a given time interval. The components of a NWS resource monitoring system are: a *name server*: a centralized controller that keeps a registry of all components and monitoring activities; *sensors* that produce resource observations; *memories* that store resource observations; and *forecasters* that process resource observations. Limits of the NWS architecture come from the presence of the name server. It introduces bottleneck, single point of failure, and does not embed security mechanism. Moreover the connections between sensors and the name server are manually managed by the administrator before starting the system. NWS is hardly applicable to a grid scale, mostly because of the absence of a real database.

2.2 Globus Monitoring and Discovery System (MDS)

The Monitoring and Discovery System [3] is the information services component of the Globus Toolkit and provides information about the available resources on the grid and their status. MDS is based on the Lightweight Directory Access Protocol (LDAP). The distributed nature of the LDAP architecture is appealing: information is organized hierarchically, and the resulting tree might be distributed over different servers. In a grid perspective, the hierarchy often reflects the organization of the grid: from continental networks to national networks to local networks. Leaves are single resource, like cluster of computers, single computer or storage element.

However, the LDAP architecture is appropriate to store static information, like the number of processors in a cluster, or the size of a disk partition. When data are more frequently changing, the LDAP architecture is a wrong choice: it is unsuitable to support frequent write operations. Indeed the LDAP Client Update Protocol is based on the assumption that "data changes, renames, and deletions of large subtrees are very infrequent" [4]. Under that condition LDAP architecture suffers serious performance and scalability problems.

2.3 Relational Grid Monitoring Architecture (R-GMA)

The Structured Query Language (SQL) allows manipulation of a relational database. Several implementations of this database are designed for extremely demanding applications. They offer a good compromise between the distributiveness, and the cost of