Lecture Notes in Computer Science 2944

Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Hong Kong
London
Milan
Paris
Tokyo

Karl Aberer Manolis Koubarakis
Vana Kalogeraki (Eds.)

Databases, Information Systems, and Peer-to-Peer Computing

First International Workshop, DBISP2P 2003
Berlin, Germany, September 7 – 8, 2003
Revised Papers

Springer

Volume Editors

Karl Aberer
Swiss Federal Institute of Technology (EPFL)
Distributed Information Systems Laboratory
School of Computer and Communication Sciences
1015 Lausanne, Switzerland
E-mail: karl.aberer@epfl.ch

Manolis Koubarakis
Technical University of Crete
Dept. of Electronic and Computer Engineering
Chania 73100 Crete, Greece
E-mail: manolis@intelligence.tuc.gr

Vana Kalogeraki
University of California, Riverside
Dept. of Computer Science and Engineering
Riverside, CA 92521, USA
E-mail: vana@cs.ucr.edu

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

CR Subject Classification (1998): H.2, H.3, H.4, C.2, I.2.11, D.2.12, D.4.3, E.1

ISSN 0302-9743
ISBN 3-540-20968-9 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2004
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Protago-TeX-Production GmbH
Printed on acid-free paper SPIN: 10982693 06/3142 5 4 3 2 1 0

Preface

Peer-to-peer (P2P) computing is currently attracting enormous media attention, spurred by the popularity of file sharing systems such as Napster, Gnutella and Morpheus. In P2P systems a very large number of autonomous computing nodes (the peers) pool together their resources and rely on each other for data and services.

The wealth of business opportunities promised by P2P networks has generated much industrial interest recently, and has resulted in the creation of various industrial projects, startup companies, and special interest groups. Researchers from distributed computing, networks, agents and databases have also become excited about the P2P vision, and papers tackling open problems in this area have started appearing in high-quality conferences and workshops.

Much of the recent research on P2P systems seems to be carried out by research groups with a primary interest in distributed computation and networks. This workshop concentrated on the impact that current database research can have on P2P computing and vice versa. Although researchers in distributed data structures and databases have been working on related issues for a long time, the developed techniques are simply not adequate for the new paradigm. P2P computing introduces the paradigm of decentralization going hand in hand with an increasing self-organization of highly autonomous peers, thus departing from the classical client-server computing paradigm. This new paradigm bears the potential to realize computing systems that scale to very large numbers of participating nodes. Taking advantage of this potential for the area of data management is a challenge that the database community itself is asked to face. The realization of the P2P computing vision is however a Herculean task, fraught with immense technical difficulties. As a result, it offers database theoreticians and system developers a new set of exciting open problems.

We believe that database research has much to contribute to the P2P grand challenge through its wealth of techniques for sophisticated semantics-based data models, clever indexing algorithms and efficient data placement, query processing techniques and transaction processing. The database community could benefit from the P2P computing vision by developing loosely coupled federations of databases where databases can join and leave the network at will; a single global schema is not a possibility, and answers need not be complete but should be best effort.

Database technologies in the new information age will form the crucial components of the first generation of complex adaptive information systems. These are an emerging kind of information systems that are very dynamic, self-organize continously and adapt to new circumstances, they are locally but not globally optimized, and form a whole which is greater than the sum of its parts. These new information systems support highly dynamic, ever-changing, autonomous social organizations and can no longer be developed using traditional analy-

sis, design and implementation techniques. This workshop also concentrated on complex adaptive information systems, their impact on current database technologies and their relation to emerging industrial technologies such as IBM's autonomic computing initiative.

This workshop, collocated with VLDB, the major international database and information systems conference, brought together key researchers from all over the world working on databases and P2P computing with the intention of strengthening this connection. Also researchers from other related areas such as distributed systems, networks, multiagent systems and complex systems participated.

The workshop was jointly organized with the AP2PC workshop which is part of the AAMAS conference and is under the responsibility of the same steering committee. Together these two workshops address both the agent and the database communities and thus take account of the interdisciplinary nature of P2P computing.

The DBISP2P workshop received 32 submissions that entered the review process. All submissions underwent a rigorous review that was completed by an online PC discussion for making the final selection of 16 papers. The organizers would like to thank at this point all program committee members for their excellent work. The program was completed by a keynote speech and a panel. The keynote speech with the title "Design Issues and Challenges for RDF- and Schema-Based Peer-to-Peer Systems" was presented by Wolfgang Nejdl from the University of Hannover. Aris Ouksel organized a panel on the topic "P2P Computing and Database Technologies: Convergence of Technologies and Socioeconomic Characteristics on the Web, Benefits and Technical Challenges in Database Applications" with the goal to explore the promise of P2P to offer exciting new possibilities in distributed information processing.

The organizers would particularly like to thank Klemens Böhm from the University of Magdeburg for his excellent work in taking care of the local arrangements, the VLDB organization for their valuable support of the workshop organization, and the steering committee for the opportunity to set up this workshop and for their continuing support.

September 2003 Karl Aberer, Manolis Koubarakis, Vana Kalogeraki

Organization

The 1st International Workshop on Databases, Information Systems and Peer-to-Peer Computing took place at Humboldt University, Berlin, Germany on September 7–8, 2003, collocated with VLDB 2003.

Workshop Chairs

Program Chairs	Karl Aberer (EPFL, Lausanne, Switzerland) Manolis Koubarakis (Technical University of Crete, Greece) Vana Kalogeraki (University of California, Riverside, USA)
Panel Chair	Aris Ouksel (University of Illinois, Chicago, USA)
Organization Chair	Klemens Böhm (University of Magdeburg, Germany)

Steering Committee

Karl Aberer (EPFL, Lausanne, Switzerland)
Sonia Bergamaschi (University of Modena and Reggio-Emilia, Italy)
Manolis Koubarakis (Technical University of Crete)
Paul Marrow (BTexact Technologies,UK)
Gianluca Moro (University of Bologna, Cesena, Italy)
Aris M. Ouksel (University of Illinois, Chicago, USA)
Claudio Sartori (CNR-CSITE, University of Bologna, Italy)
Munindar P. Singh (North Carolina State University, USA)

Program Committee

Ozalp Babaoglu, University of Bologna, Italy
Klemens Böhm, University of Magdeburg, Germany
Beng Chin Ooi, National University of Singapore, Singapore
Partha Dasgupta, Arizona State University, USA
Alex Delis, Polytechnic University, USA
Fausto Giunchiglia, University of Trento, Italy
Zachary G. Ives, University of Washington, USA
Carole Goble, University of Manchester, UK
Oliver Guenther, Humboldt University, Germany
Dimitris Gunopoulos, University of California at Riverside, USA
Manfred Hauswirth, EPFL, Switzerland
Achilles D. Kameas, Computer Technology Institute, Greece
Yannis Labrou, Fujitsu Labs of America, USA
Witold Litwin, University of Paris 6, France
Ling Liu, Georgia Institute of Technology, USA
Peri Loucopoulos, UMIST, Manchester, UK
Dejan Milojicic, Hewlett-Packard Labs, USA
Alberto Montresor, University of Bologna, Italy
Jean-Henry Morin, University of Geneva, Switzerland
John Mylopoulos, University of Toronto, Canada
Wolfgang Nejdl, Learning Lab Lower Saxony, Germany
Dimitris Papadias, Hong Kong University of Science and Technology, China
Mike Papazoglou, Tilburg University, Netherlands
Evaggelia Pitoura, University of Ioannina, Greece
Dimitris Plexousakis, Institute of Computer Science, FORTH, Greece
Onn Shehory, IBM Haifa, Israel
Spiros Skiadopoulos, National Technical University of Athens, Greece
Katia Sycara, Robotics Institute, Carnegie Mellon University, USA
Peter Triantafillou, University of Patras, Greece
Martin Wolpers, Learning Lab Lower Saxony, Germany

Referees

W. Black
Claus Boyens
David Buttler
James Caverlee
Eleni Christopoulou
Matthias Fischmann
Christos Goumopoulos
Tasos Gounaris
Verena Kantere
Mujtaba Khambatti
Georgia Koloniari
Bin Liu
Nikolaos Ntarmos
Themis Palpanas
Michel Pawlak
I. Petrounias
Theoni Pitoura
Th. Schwarz

Table of Contents

Invited Talk

Design Issues and Challenges for RDF- and Schema-Based Peer-to-Peer Systems 1
Wolfgang Nejdl (Learning Lab Lower Saxony and University of Hannover)

Structure in P2P Networks

SIL: Modeling and Measuring Scalable Peer-to-Peer Search Networks 2
Brian F. Cooper and Hector Garcia-Molina (Stanford University)

Searchable Querical Data Networks 17
Farnoush Banaei-Kashani and Cyrus Shahabi (University of Southern California)

Semantic Overlay Clusters within Super-Peer Networks 33
Alexander Löser (Technische Universität Berlin), Felix Naumann (Humboldt University Berlin), Wolf Siberski, Wolfgang Nejdl, and Uwe Thaden (Learning Lab Lower Saxony)

Structuring Peer-to-Peer Networks Using Interest-Based Communities . . . 48
Mujtaba Khambatti, Kyung Dong Ryu, and Partha Dasgupta (Arizona State University)

Semantics and Data Integration

A Robust Logical and Computational Characterization for Peer-to-Peer Database Systems 64
Enrico Franconi (Free University of Bozen-Bolzano, Italy), Gabriel Kuper (University of Trento, Italy), Andrei Lopatenko (Free University of Bozen-Bolzano, Italy, University of Manchester, UK), and Luciano Serafini (ITC-irst, Trento, Italy)

Semantic Data Integration in P2P Systems 77
Diego Calvanese, Elio Damaggio, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati (Università di Roma "La Sapienza")

Defining Peer-to-Peer Data Integration Using Both as View Rules 91
Peter McBrien (Imperial College London) and Alexandra Poulovassilis (University of London)

Coordinating Peer Databases Using ECA Rules 108
Vasiliki Kantere (University of Toronto, Canada), Iluju Kiringa (University of Ottawa, Canada), John Mylopoulos, Anastasios Kementsietsidis, and Marcelo Arenas (University of Toronto, Canada)

Data Streams and Publish/Subscribe

An Adaptive and Scalable Middleware for Distributed Indexing of Data Streams 123
Ahmet Bulut, Roman Vitenberg, Fatih Emekçi, and Ambuj K. Singh (UCSB, Santa Barbara)

Building Content-Based Publish/Subscribe Systems with Distributed Hash Tables 138
David Tam, Reza Azimi, and Hans-Arno Jacobsen (University of Toronto)

Data Structures and Query Processing

AmbientDB: Relational Query Processing in a P2P Network 153
Peter Boncz and Caspar Treijtel (CWI)

Towards a Unifying Framework for Complex Query Processing over Structured Peer-to-Peer Data Networks 169
Peter Triantafillou and Theoni Pitoura (University of Patras, Greece)

Distributed Queries and Query Optimization in Schema-Based P2P-Systems 184
Ingo Brunkhorst (Learning Lab, Lower Saxony, Germany), Hadami Dhraief (University of Hannover, Germany), Alfons Kemper (University of Passau, Germany), Wolfgang Nejdl (Learning Lab, Lower Saxony and University of Hannover, Germany), and Christian Wiesner (University of Passau, Germany)

PePeR: A Distributed Range Addressing Space for Peer-to-Peer Systems 200
Antonios Daskos, Shahram Ghandeharizadeh, and Xinghua An (University of Southern California)

Efficient Search in Structured Peer-to-Peer Systems: Binary v.s. k-Ary Unbalanced Tree Structures 219
Magdalena Punceva and Karl Aberer (EPFL, Switzerland)

Content-Based Overlay Networks for XML Peers Based on Multi-level Bloom Filters 232
Georgia Koloniari, Yannis Petrakis, and Evaggelia Pitoura (University of Ioannina, Greece)

Author Index 249