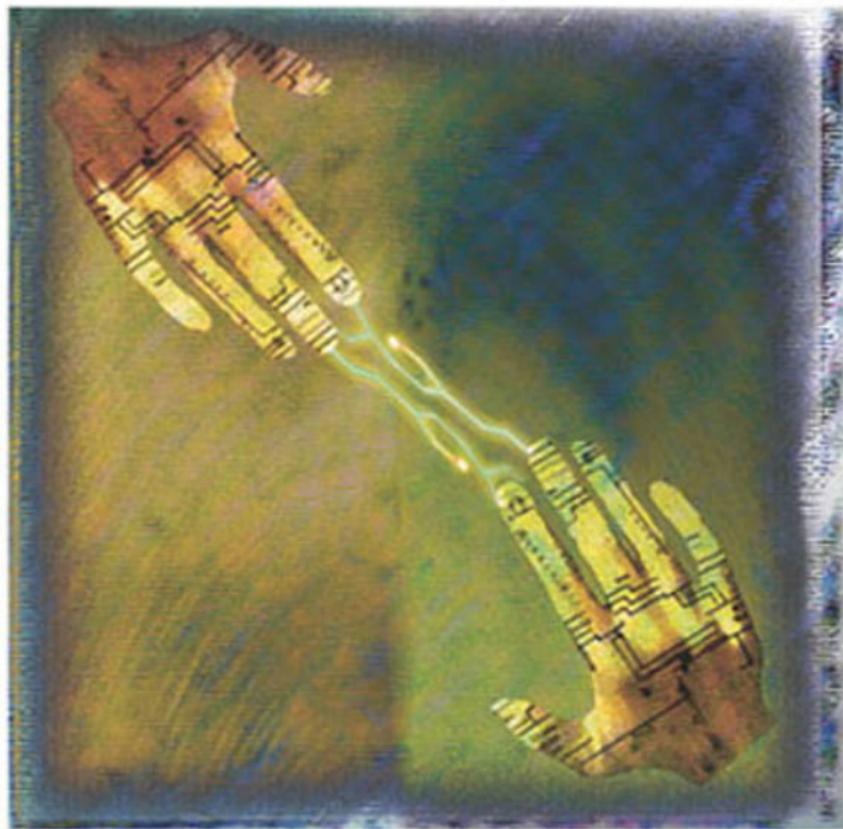


PEER^{TO}PEER COMPUTING

The Evolution of a Disruptive Technology



Ramesh Subramanian
& Brian D. Goodman

Peer-to-Peer Computing: The Evolution of a Disruptive Technology

Ramesh Subramanian
Quinnipiac University, USA

Brian D. Goodman
IBM Corporation, USA



IDEA GROUP PUBLISHING

Hershey • London • Melbourne • Singapore

Acquisitions Editor: Mehdi Khosrow-Pour
Senior Managing Editor: Jan Travers
Managing Editor: Amanda Appicello
Development Editor: Michele Rossi
Copy Editor: Joyce Li
Typesetter: Sara Reed
Cover Design: Lisa Tosheff
Printed at: Integrated Book Technology

Published in the United States of America by
Idea Group Publishing (an imprint of Idea Group Inc.)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@idea-group.com
Web site: <http://www.idea-group.com>

and in the United Kingdom by
Idea Group Publishing (an imprint of Idea Group Inc.)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 3313
Web site: <http://www.eurospan.co.uk>

Copyright © 2005 by Idea Group Inc. All rights reserved. No part of this book may be reproduced in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher.

Library of Congress Cataloging-in-Publication Data

Peer-to-peer computing : the evolution of a disruptive technology / Ramesh Subramanian and Brian D. Goodman, editors.

p. cm.

Includes bibliographical references and index.

ISBN 1-59140-429-0 (hard cover) -- ISBN 1-59140-430-4 (soft cover) -- ISBN 1-59140-431-2 (Ebook)

1. Peer-to-peer architecture (Computer networks) I. Subramanian, Ramesh. II. Goodman, Brian D.

TK5105.525.P443 2004

004.6'5--dc22

2004022155

British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Peer-to-Peer Computing: The Evolution of a Disruptive Technology

Table of Contents

Prefaceix

Section I: Then and Now: Understanding P2P Spirit, Networks, Content Distribution and Data Storage

Chapter I

Core Concepts in Peer-to-Peer Networking 1

Detlef Schoder, University of Cologne, Germany

Kai Fischbach, University of Cologne, Germany

Christian Schmitt, University of Cologne, Germany

Chapter II

Peer-to-Peer Networks for Content Sharing28

Choon Hoong Ding, The University of Melbourne, Australia

Sarana Nutanong, The University of Melbourne, Australia

Rajkumar Buyya, The University of Melbourne, Australia

Chapter III

Using Peer-to-Peer Systems for Data Management66

Dinesh C. Verma, IBM T.J. Watson Research Center, USA

Chapter IV

Peer-to-Peer Information Storage and Discovery Systems79

Cristina Schmidt, Rutgers University, USA

Manish Parashar, Rutgers University, USA

Section II: Systems and Assets: Issues Arising from Decentralized Networks in Security and Law

Chapter V

Peer-to-Peer Security Issues in Nomadic Networks 114

Ross Lee Graham, Mid-Sweden University, ITM, Sweden

Chapter VI

Potential Security Issues in a Peer-to-Peer Network from a Database Perspective 131

Sridhar Asvathanarayanan, Quinnipiac University, USA

Chapter VII

Security and Trust in P2P Systems 145

Michael Bursell, Cryptomathic, UK

Chapter VIII

Peer-to-Peer Technology and the Copyright Crossroads..... 166

Stacey L. Dogan, Northeastern University School of Law, USA

Section III: P2P Domain Proliferation: Perspectives and Influences of Peer Concepts on Collaboration, Web Services and Grid Computing

Chapter IX

Personal Peer-to-Peer Collaboration Based on Shared Objects 195

Werner Geyer, IBM T.J. Watson Research Center, USA

Juergen Vogel, University of Mannheim, Germany

Li-Te Cheng, IBM T.J. Watson Research Center, USA

Michael J. Muller, IBM T.J. Watson Research Center, USA

Chapter X

“Let Me Know What You Know”: ReachOut as a Model for a P2P Knowledge Sharing Network 225

Vladimir Soroka, IBM Haifa Research Lab, Israel

Michal Jacovi, IBM Haifa Research Lab, Israel

Yoelle S. Maarek, IBM Haifa Research Lab, Israel

Chapter XI

Ten Lessons from Finance for Commercial Sharing of IT Resources 244

Giorgos Cheliotis, IBM Research GmbH, Switzerland

Chris Kenyon, IBM Research GmbH, Switzerland

Rajkumar Buyya, University of Melbourne, Australia

Chapter XII	
Applications of Web Services in Bioinformatics	265
<i>Xin Li, University of Maryland Baltimore, USA</i>	
<i>Aryya Gangopadhyay, University of Maryland Baltimore, USA</i>	
Chapter XIII	
Content Delivery Services in a Grid Environment	278
<i>Irwin Boutboul, IBM Corporation, USA</i>	
<i>Dikran S. Meliksetian, IBM Corporation, USA</i>	
About the Editors	296
About the Authors	298
Index	305

Foreword

After decades of growth, we are now about 5% of the way into what the Internet has in store for our business and personal lives. Soon, a billion people will be using the Net, empowering themselves to get what they want, when they want it, from wherever they are. Each day we get closer to a new phase of the Internet that will make today's version seem primitive. Not only will this next-generation Internet be orders of magnitude faster, but it also will be always on, everywhere, natural, intelligent, easy, and trusted.

Fast and reliable connectivity is finally appearing and the competition to provide it is beginning to heat up. Cable, telecom, satellite, and the power grid are each threatening the other and the result will be more speed, improved service, and lower prices. More important than the speed is the always-on connection, which will increase propensities to use online services—and also increase expectations. The impact of WiFi is bigger than coffee shops and train stations. With WiFi chips in handheld devices and the rapid adoption of voice over IP, the Internet becomes available everywhere and a voice conversation becomes just one of the many things you can do while connected. Long distance will no longer mean anything. WiFi will soon be as secure and as fast as today's wired Ethernet. Advanced antenna and radio technologies will ensure ubiquity. With more people always on and having adequate bandwidth, information-oriented e-businesses will lead the charge for the reemergence of the application service provider.

Web services are enabling a global application Web where any and all applications can be linked together seamlessly. Not only will you be able to use frequent flyer points to pay for hotel reservations online, but also to designate from a checkbox on that same hotel Web page the airline from whose frequent-flier program the points should be deducted.

It will soon be clear that Linux is not about “free.” It is about achieving scalability, reliability, and security. The world will remain heterogeneous but the underlying operating systems need to be open so that all can see how it works and contribute to it. The “open source” model also will mean more rapid innovation.

Security will no longer be the biggest issue—authentication will. Digital certificates will enable people, computers, handhelds, and applications to interact se-

curely in a distributed Web of trust. With a redesign of e-mail protocols, we also will gain confidence and control over whom we communicate with.

The potential of the Internet is much greater than meets the eye. As the Internet evolves, it will become so pervasive, reliable, and transparent that we will take it for granted. It will be part of our life and, more important, begin to simplify our lives.

One of the many magical elements of the Internet is that every computer connected to it is also connected to every other computer connected to it. There is no central switching office as with the telephone system. Some of the computers on the Net are servers providing huge amounts of information and transactions, but most of the computers are home and office PCs operated by individuals. When one of these individuals connects with another one, it is called a peer-to-peer connection.

Like most technologies that have gained attention on the Internet, peer-to-peer is not a new idea. Peer-to-peer went mainstream during the dot com era of the late 1990s when a teenager named Shawn Fenning appeared on the cover of *Time* magazine after having founded a company called Napster. Napster devised a technology for using peer-to-peer connections to exchange compressed music files (MP3s). Because MP3 music downloaded from the Net sounds the same as music from a CD, and because there are millions of college students with fast Internet connections, the peer-to-peer phenomenon experienced a meteoric growth in popularity.

The recording industry should have anticipated music sharing but instead found itself on the defense and then resorted to legal action to stem the tide. Over the next few years, we will find out if it was too late and the upstarts such as tunes will reshape the music industry.

But peer-to-peer is much bigger than music sharing. It is also information sharing. Not just college students but also business colleagues. Not just music but video conferences. Not just for fun but for serious collaboration in business, government, medicine, and academia. Not just person to person but peer-to-peer networks of many persons—millions, perhaps hundreds of millions. Not just communicating and sharing but combining the computing power of large numbers of computers to find life in outer space, a cure for cancer, or how to untangle the human genome.

It is understandable that the music industry committed itself to an all-out fight against the explosion of peer-to-peer file sharing networks. It is also understandable that many major enterprises have banned peer-to-peer file sharing tools because of a concern that their employees may be importing illegally obtained intellectual property and also out of a justified fear that peer-to-peer networks have spread deadly viruses.

Peer-to-peer is too important to be categorically banned. It needs to be understood and exploited for its merits while policy makers work through the legal

and societal issues. Once we truly understand peer-to-peer, we will find that the reality exceeds the hype.

Peer-to-Peer computing: The Evolution of a Disruptive Technology is an important book because it unravels the details of peer-to-peer. This cohesive body of work focuses on the genesis of peer-to-peer—the technologies it is based on, its growth, its adoption in various application areas, and its economic and legal aspects. It also goes deep into peer-to-peer across a broad range of technologies including file sharing, e-mail, grid-based computing, collaborative computing, digital asset management, virtual organizations, new ways of doing business, and the legal implications.

Subramanian and Goodman combine their academic and technology talents to create a compendium filled with practical ideas from existing projects. The book offers a view of peer-to-peer through a series of current articles from academics, IT practitioners, and consultants from around the world.

If you are interested in a complete picture of peer-to-peer technologies, their foundations and development over the years, their applications and business and commercial aspects, then this is a great reference text. Whether you want to gain a basic understanding of peer-to-peer or dive deep into the complex technical aspects, you will find this book a great way to gain ideas into the future of peer-to-peer computing.

John R. Patrick
President, Attitude LLC
Connecticut
May 2004

Preface

In May 1999, Shawn Fanning and Sean Parker created Napster Inc., thus beginning an unforeseen revolution. At the time, Napster was arguably the most controversial free peer-to-peer (P2P) file sharing system the Internet had ever seen. Napster was in many ways an expression of the underground movement that came before it—the world of bulletin board systems, anonymous FTP servers, and the idea of *warez*. *Warez* refers to pirated software that has been modified or packaged with registration information. Anyone in possession of *warez* is able to install and run the software as if they had purchased the real license. The successful propagation of pirated software on the Internet is directly attributable to the ease with which loosely associated but highly organized communities can be formed and maintained on the Net. Napster not only answered the need for an easy way to find and share music files, but it also built a community around that concept. People make copies of video, audiotapes, and CDs for personal use all the time. They sometimes share these copies with other people as simply *part of their social mores*. The advent of the MP3 audio format has made the exchange of music all the more easy. People can quickly digitize their music collections and share them with others, using the Internet. Indeed, the Internet provides an extraordinary ability to abuse copyright; it is fast, relatively easy, and with the entry of file sharing software, music can be shared with not just one friend, but with anybody in the world who desires it.

Let's fast-forward to the present time. Now, after endless litigation spearheaded by the Recording Industry Association of America (RIAA), Napster is a for-profit business with strong ties to the music trade—a different avatar from its original revolutionary self.

Chronologies of P2P computing often begin with a reference to Napster. It is the most popular example of just how powerfully one-to-one and one-to-many communications can be realized through computing technology. However, if we look further back, instant messaging was probably an earlier incarnation of P2P. Instant messaging represents a different form of communication. People no longer write as many e-mails—they are engaging in real-time messaging.

Instant messaging provides a compelling hybrid of the telephone and letter writing; all the immediacy of a phone call with all the control of an e-mail. Instant messaging has transformed the Internet landscape and continues to revolutionize the business world.

In fact, from a technology viewpoint, peer-to-peer computing is one of those revisits to past technologies and mind-sets. Often, really great ideas are initially met with little embrace as the environment in which they might flourish lacks nourishment. The concepts that made Napster a reality are not new. Napster simply became an icon of the great P2P underground movement by bringing to reality some of the most basic networking concepts that have existed for a long time. Napster's success was shared by other similar, contemporaneous tools, and the buzz this generated underscored the fact that the time was indeed right for a technology revisit.

P2P computing has become so commonplace now that some regard it as old news. However, the reality is that we have yet to discover all the ramifications of P2P computing—the maturity of peer systems, the proliferation of P2P applications, and the continually evolving P2P concepts are new.

The goal of this book is to provide insight into this continuing evolution of P2P computing more than four years after its popular and notorious debut. It draws upon recent relevant research from both academia and industry to help the reader understand the concepts, evolution, breadth, and influence of P2P technologies and the impact that these technologies have had on the IT world. In order to explore the evolution of P2P as a disruptive technology, this book has been broken up into three major sections. Section I begins by exploring some of P2P's past—the basic underpinnings, the networks, and the direction they began to take as distribution and data systems. Section II addresses trust, security and law in P2P systems and communities. Section III explores P2P's domain proliferation. It attempts to capture some of the areas that have been irreversibly influenced by P2P approaches, specifically in the area of collaboration, Web services, and grid computing.

Looking at Disruptive Technologies

Disruptive technologies are at the heart of change in research and industry. The obvious challenge is to distinguish the hype from reality. Gartner Research's "Hype Cycles" work (see Figure 1) charts technologies along a life-cycle path, identifying when the technology is just a buzzword through to its late maturation or productivity (Linden and Fenn, 2001, 2003). In 2002, peer-to-peer computing was entering the *Trough of Disillusionment*. This part of the curve represents the technologies' failure to meet the hyped expectations. Every technology en-

Figure 1. Hype cycles



Source: Gartner Research (May 2003)

ters this stage where activities in the space are less visible. Business and venture capitalists continue to spend time and money as the movement climbs the *Slope of Enlightenment* beginning the path of adoption. It is thought that peer-to-peer will plateau anywhere from the year 2007 to 2012. As the peer-to-peer mind-set continues to permeate and flourish across industries, there is a greater need to take a careful reading of the technology pulse. Peer-to-peer represents more than file sharing and decentralized networks. This book is a collection of chapters exemplifying cross-domain P2P proliferation—a check of the P2P pulse.

The Book

Section I of the book deals with the issues of “then and now”—understanding P2P spirit, networks, content distribution, and data storage.

In Chapter I, Detlef Schoder, Kai Fischbach, and Christian Schmitt review the core concepts in peer-to-peer networking. Some of the issues that the authors address are the management of resources such as bandwidth, storage, information, files, and processor cycles using P2P networks. They introduce a model that differentiates P2P infrastructures, P2P applications, and P2P communities. Schoder et al. also address some of the main technical as well as social chal-

lenges that need to be overcome in order to make the use of P2P more widespread.

Choon Hoong Ding, Sarana Nutanong, and Rajkumar Buyya continue the overview of P2P computing in Chapter II with a special focus on network topologies used in popular P2P systems. The authors identify and describe P2P architectural models and provide a comparison of four popular file sharing software—namely, Napster, Gnutella, Fasttrack, and OpenFT.

Historically, most peer-to-peer work is done in the area of data sharing and storage. Chapter III focuses on modern methods and systems addressing data management issues in organizations. Dinesh Verma focuses on the data storage problem and describes a peer-to-peer approach for managing data backup and recovery in an enterprise environment. Verma argues that data management systems in enterprises constitute a significant portion of the total cost of management. The maintenance of a large dedicated backup server for data management requires a highly scalable network and storage infrastructure, leading to a major expense. Verma suggests that an alternative peer-to-peer paradigm for data management can provide an approach that provides equivalent performance at a fraction of the cost of the centralized backup system.

Continuing the theme of data storage, Cristina Schmidt and Manish Parashar investigate peer-to-peer (P2P) storage and discovery systems in Chapter IV. They present classification of existing P2P discovery systems, the advantages and disadvantages of each category, and survey existing systems in each class. They then describe the design, operation, and applications of Squid, a P2P information discovery system that supports flexible queries with search guarantees.

Section II of the book shifts the focus to systems and assets, and the issues arising from decentralized networks in diverse areas such as security and law. In Chapter V, Ross Lee Graham traces the history of peered, distributed networks, and focuses on their taxonomy. He then introduces nomadic networks as implementations of peer-to-peer networks, and discusses the security issues in such networks, and then provides a discussion on security policies that could be adopted with a view to building trust management.

Sridhar Asvathanarayanan takes a data-centered approach in Chapter VI, and details some of the security issues associated with databases in peer networks. Microsoft Windows® is currently one of the most popular operating systems in the world and in turn is a common target environment for peer-to-peer applications, services, and security threats. Asvathanarayanan uses Microsoft® SQL server as an example to discuss the security issues involved in extracting sensitive data through ODBC (open database connectivity) messages and suggests ways in which the process could be rendered more secure. The author underscores that security starts by analyzing and being treated at the technology level.

Michael Bursell offers a more holistic focus on security in Chapter VII by examining the issue of security in peer-to-peer (P2P) systems from the standpoint of trust. The author defines trust, explains why it matters and argues that trust as a social phenomenon. Taking this socio-technical systems view, the author identifies and discusses three key areas of importance related to trust: identity, social contexts, and punishment and deterrence. A better understanding of these areas and the trade-offs associated with them can help in the design, implementation, and running of P2P systems.

In Chapter VIII, law professor Stacey Dogan discusses the challenges that peer-to-peer networks pose to the legal and economic framework of United States Copyright Law. According to Dogan, peer-to-peer networks “debunk the historical assumption that copyright holders could capture their core markets by insisting on licenses from commercial copiers and distributors who actively handled their content.” The main way by which peer-to-peer networks accomplish that is through the adherence to communitarian values such as sharing and trust. In this chapter, the author explains why peer-to-peer technology presents such a challenge for copyright, and explores some of the pending proposals to solve the current dilemma.

After addressing the complex and seemingly intractable issues such as security and law as they relate to peer-to-peer networks, we move to Section III of the book, which deals with P2P domain proliferation—the applications of peer-to-peer computing, and the perspectives and influences of peer concepts in the areas of collaboration, Web services, and grid computing.

Peer-to-peer computing has been promoted especially by academics and practitioners alike as the next paradigm in person-to-person collaboration. In Chapter IX, Werner Geyer, Juergen Vogel, Li-Te Cheng, and Michael Muller describe the design and system architecture of such a system that could be used for personal collaboration. Their system uses the notion of shared objects such as a chat mechanism and a shared whiteboard that allow users to collaborate in a rich but lightweight manner. This is achieved by organizing different types of shared artifacts into semistructured activities with dynamic membership, hierarchical object relationships, and synchronous and asynchronous collaboration. The authors present the design of a prototype system and then develop an enhanced consistency control algorithm that is tailored to the needs of this new environment. Finally, they demonstrate the performance of this approach through simulation results.

In Chapter X, Vladimir Soroka, Michal Jacovi, and Yoelle Maarek continue the thread on P2P collaboration and analyze the characteristics that make a system peer-to-peer and offer a P2P litmus test. The authors classify P2P knowledge sharing and collaboration models and propose a framework for a peer-to-peer systems implementation that is an advancement over existing models. They refer to this model as the *second degree peer-to-peer model*, and illustrate it with ReachOut, a tool for peer support and community building.

In Chapter XI, Giorgos Cheliotis, Chris Kenyon, and Rajkumar Buyya introduce a new angle to the discussion of P2P applications and implementations. They argue that even though several technical approaches to resource sharing through peer-to-peer computing have been established, in practice, sharing is still at a rudimentary stage, and the commercial adoption of P2P technologies is slow because the existing technologies do not help an organization decide how best to allocate its resources. They compare this situation with financial and commodity markets, which “have proved very successful at dynamic allocation of different resource types to many different organizations.” Therefore they propose that the lessons learned from finance could be applied to P2P implementations. They present 10 basic lessons for resource sharing derived from a financial perspective and modify them by considering the nature and context of IT resources.

In Chapter XII, Xin Li and Aryya Gangopadhyay introduce applications of Web services in bioinformatics as a specialized application of peer-to-peer (P2P) computing. They explain the relationship between P2P and applications of Web services in bioinformatics, state some problems faced in current bioinformatics tools, and describe the mechanism of Web services framework. The authors then argue that a Web services framework can help to address those problems and give a methodology to solve the problems in terms of composition, integration, automation, and discovery.

In Chapter 13, Irwin Boutboul and Dikran Meliksetian describe a method for content delivery within a computational grid environment. They state that the increasing use of online rich-media content, such as audio and video, has created new stress points in the areas of content delivery. Similarly, the increasing size of software packages puts more stress on content delivery networks. New applications are emerging in such fields as bio-informatics and the life sciences that have increasingly larger requirements for data. In parallel, to the increasing size of the data sets, the expectations of end users for shorter response times and better on-demand services are becoming more stringent. Moreover, content delivery requires strict security, integrity, and access control measures.

All those requirements create bottlenecks in content delivery networks and lead to the requirements for expensive delivery centers. The authors argue that the technologies that have been developed to support data retrieval from networks are becoming obsolete, and propose a grid-based approach that builds upon both grid technologies and P2P to solve the content delivery issue. This brings us full circle and exemplifies how at the core of content distribution lies a discernible P2P flavor.

References

- Linden, A., & Fenn, J. (2001). 2002 emerging technologies hype cycle: Trigger to peak. Gartner, COM-16-3485, 2.
- Linden, A., & Fenn, J. (2003). Understanding Gartner's hype cycles. Gartner, R-20-1971.

Acknowledgments

We would like to thank all the authors for their diverse and excellent contributions to this book. We would also like to thank all of our reviewers and the editors and support staff at the Idea Group Inc.. We would especially like to thank our institutions—IBM Corporation and Quinnipiac University—for their generous understanding and support in providing the time and resources to bring such diverse worlds together. Finally, we would like to express our deepest gratitude to our families, friends, and colleagues for all their support during this project.

Ramesh Subramanian
North Haven, Connecticut

Brian D. Goodman
New York, New York

May 2004

Section I

Then and Now:
Understanding
P2P Spirit, Networks,
Content Distribution
and Data Storage

Chapter I

Core Concepts in Peer-to-Peer Networking

Detlef Schoder

University of Cologne, Germany

Kai Fischbach

University of Cologne, Germany

Christian Schmitt

University of Cologne, Germany

Abstract

This chapter reviews core concepts of peer-to-peer (P2P) networking. It highlights the management of resources, such as bandwidth, storage, information, files, and processor cycles based on P2P networks. A model differentiating P2P infrastructures, P2P applications, and P2P communities is introduced. This model provides a better understanding of the different perspectives of P2P. Key technical and social challenges that still limit the potential of information systems based on P2P architectures are discussed.