

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.

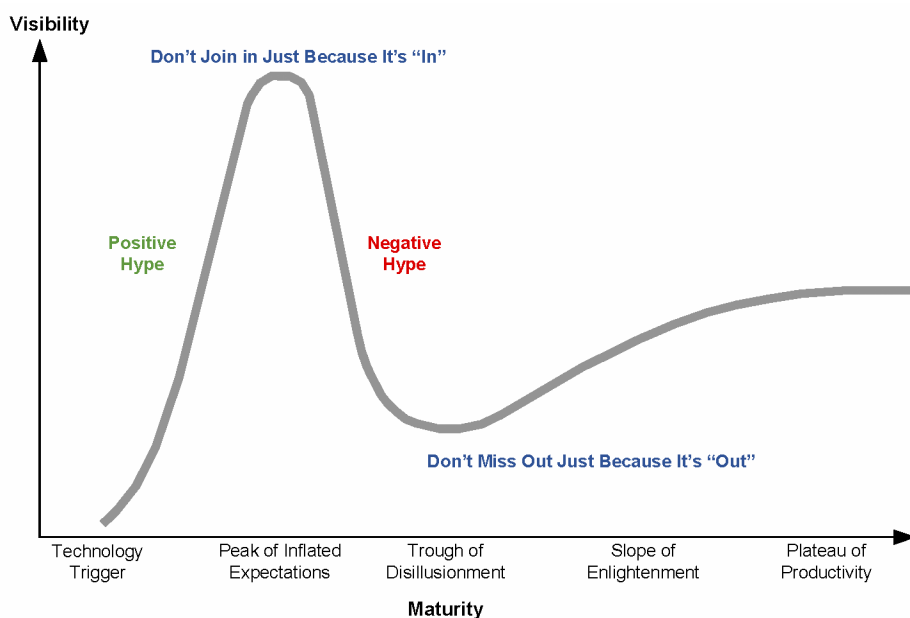
P2P computing has become so commonplace now that some regard it as old news. However, the reality is that we have yet to discover all the ramifications of P2P computing—the maturity of peer systems, the proliferation of P2P applications, and the continually evolving P2P concepts are new.

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.