Peer-to-Peer: Harnessing the Power of Disruptive Technologies

by Andy Oram (Ed.)

O'Reilly, 2001
448 pages, list price $ 29.95 (US)
ISBN 0-596-00110-X

Review by:
Mario A. Nascimento, Univ. of Alberta, Canada
mn@cs.ualberta.ca

This is a book about peer-to-peer (P2P) issues in general – and it is a good one for that matter – but the potential reader should keep in mind that it is *not* a book about data management within a P2P environment. Why it is being reviewed here, you may ask? Besides being on the list of books to be reviewed, the main reason is that this is an excellent book for those who may be considering working on P2P related problems. Andy Oram, the book's editor, did an excellent job in putting together the 19 chapters that form this book. They cover a broad range of topics centered on the P2P theme, and are divided in three main areas: context and overview, projects and technical issues. NAPSTER and file sharing in general, copyright infringement are all issues commonly associated with P2P, but P2P is much more than that. The book's great contribution is to present and discuss, using plain English, a number of less popular (abusing the term) or overseen issues related to P2P in general.

This is a book that can be read at leisure. The chapters are not dependent of each other, although they flow well in the presented order. It is not a research oriented book, instead it is a book one can use to get closer to the origins of P2P, how it has evolved, how some mainstream projects have dealt with some interesting problems, and so on. Bottomline, if the reader is really interested about the P2P domain (and not only about file sharing) he/she will likely find the domain even more interesting after reading this book.

The first part of the book lays down the book's context of P2P. Chapter 1 traces the origin of P2P back to ARPANET and USENET (and UUCP!). Whereas those had no hierarchy and could be considered grandfathers to Gnutella (as the author puts it), DNS service could be an ancestor to Napster. (Interesting enough, Napster, who popularized – maybe in a not so positive way – the term P2P, is not a strict P2P service since it relies on a centralized server.) Nonetheless the DNS scalability offers useful lessons to P2P architectures. Chapter 2 draws on the fact that decentralized P2P services, using Napster's phenomenon as an example, should be seen as a tool not a goal. It also presents an interesting discussion on the economic rather than legal challenge posed by P2P, and argues that asymmetric technologies used by ADSL/cable modem ISPs is not adequate for P2P services. Chapter 3 reports on a P2P summit held in 2000, and provides an interesting summary of opinions about the domain. The chapter is filled with personal insights, my favorite quote being one by Dave Winer: "the P in P2P is for people", which reinforces the idea of P2P as a tool not a goal. This part of the book is concluded with a short chapter with a quite intriguing view of P2P services: they do nothing much more than capitalizing on people's selfishness, i.e., the growth of the database as a whole has no altruistic reason, rather it is fed by everyone's greedy to have more.

The second part of the book talks about many P2P projects, some quite interesting but not as popular. Chapter 5 talks about SETI@home, which allows one to use underutilized CPU cycles of peers towards solving a CPU intensive specific problem. (One could argue here, that the first P in P2P can be used for Power and the second for People, but I digress.) Jabber, a very portable instant messaging service using XML as the underlying framework is presented in Chapter 6. Next, a less notorious use

of P2P is discussed, namely that one of anonymous remailers. Not as glamorous as the file sharing domain, the use of P2P to ensure freedom-of-speech and the like is convincingly argued to be an important application. Chapter 8 discusses in great detail the Gnutella network, the "internet potluck party", and a good (If not best) example of a true P2P network/framework. The chapter illustrates nicely the power behind Gnutella's protocol – a quite interesting example is one where among several "traditional" clients one was a simple calculator. If the "query" was textual the calculator would ignore it, but when it was a mathematical expression it would actually compute it and return the result, whereas the other clients would likely ignore it; all transparently to the user. The chapter, a pleasure to read, is very enlightening if one is seeking more information about Gnutella. This part of the book is concluded with Chapters 9 through 12, which present Freenet, Red Rover, Publius and Free Haven, respectively. All these, along with chapter 7, deal, in one degree or another, with the issue of anonymity, anti-censorship and document integrity, persistence and availability.

Finally, the third and last part of the book begins with a chapter dedicated to Metadata, an important issue, still largely unexplored within the P2P domain, when considering the intrinsic need to node cooperation and information sharing. Although it discusses ideas familiar to information retrieval, such as Dublin Core, and brings to the database-oriented reader's mind topics such as semantic heterogeneity and schema matching, though they are not mentioned at all. Chapter 14, about performance, is perhaps the most technical chapter in the book. It presents a fairly intuitive notion of graphs and quickly connects to the, increasingly popular, notion of "small-worlds." In a (P2P) small-world network nodes are, on average, fairly close to each other with the existence of several (clustered) communities. The chapter shows, for instance, that the Freenet network evolves naturally to a small-world network. It uses the same framework to investigate, qualitatively and quantitatively, both Freenet and Gnutella. For those more interested in P2P performance/simulation, this chapter is a must. A well designed simulating environment based on some of the simple ideas presented in that chapter could lead to more interesting results regarding, for instance, cache management. Chapter 15 discusses the notion of trust, which is quite related to the subjects in Chapters 7 and 9 through 12. (In fact, it uses the Publius system discussed in Chapter 11 as a supporting example for some of its arguments.) The authors advocate the provoking idea that the "ideal trusted system is one that everyone has confidence in because they do not have to trust it." Unfortunately, trust, as subjective as it may be, is a necessary "evil" in today's environment. Chapter 16 is the longest one in the book, and also one of the few that offers footnote references – most chapters describe the authors' experiences and insights. In this chapter authors discuss to a great extent, the problem of accountability. Issues of (micro)payment and pseudonymity are discussed and related to privacy and threat management. It also covers the issue of reputation, which is also dealt with in Chapter 17. Chapter 18 discusses the topic of security, but from the perspective of ensuring, again, mutual trust between collaborators. That differs from what one would typically expect within this topic, e.g., preventing malicious and/or denial-of-service attacks. Finally Chapter 19 concludes the book's contributions advocating for a OmniNetwork (or perhaps meta-network), achieved through gateways. In this way it might be possible to take advantage of the best features of different types of networks/services, since a super-network (or übernetwork as the author puts it) is not feasible.

In summary, the book presents and discusses a broad range of issues related to P2P; some of those will likely be addressed at some point, directly or indirectly, by the database community. If the reader is interested in data management within a P2P environment, the only resources seem to be recent conferences and workshops[1]. SIGMOD Record should feature a special issue on the topic later this year. Finally, at the time of this writing there are open CFPs for a few workshops as well as for a special issue of TKDE (forthcoming in the first half of 2004) that should be of great interest to those searching for academic material related to P2P data management.

[1] Needless to say, a web search for "peer-to-peer data management" is probably also a good idea (as of now it would reveal pointers to several research groups).