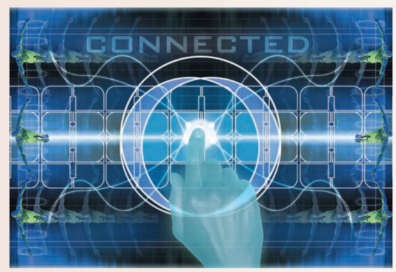
P2P Networking: An Information-Sharing Alternative

Peer-to-peer networking offers unique advantages that will make it a more effective alternative to several existing client-server e-commerce applications—if it can mature into a secure and reliable technology.



Manoj Parameswaran
University of Maryland

Anjana Susarla
Andrew B. Whinston
University of Texas at Austin

The Internet's phenomenal impact and subsequent growth stemmed primarily from the open, public, worldwide networking it offers. Most current applications, however, use client-server technologies that incorporate networking as an ancillary, value-added feature. Peer-to-peer computing offers a radically new way of isolating and focusing on the networking aspect as the business model's mainstay. A P2P network distributes information among the member nodes instead of concentrating it at a single server. This paradigm offers exciting advantages in information sharing, but it also presents challenging disadvantages.

P2P networks differ markedly from the client-server interactions that typify applications in the TCP/IP world. A client-server scenario like the Web depends on a single server storing information and distributing it to clients in response to their requests. The information repository remains essentially static, centralized at the server, and subject only to updates by the provider. Users assume a passive role in that they receive, but do not contribute, information. A P2P network, on the other hand, considers all nodes equal in their capacity for sharing information with other network members. Each user makes an information repository available for distribution, which, combined with anyone's ability to join the network, leads to the fast growth of a network composed of distributed information repositories.

In a P2P model, each member node can make information available for distribution and can establish direct connections with any other member node to download information. Instead of looking at what is available in a centralized repository, such as MP3.com, a client seeking information from a P2P network searches across scattered collections stored at numerous member nodes, all of which appear to be a single repository with a single index, such as Napster.

Member nodes primarily use existing P2P networks, such as Gnutella or Napster, to share information of some kind. The TCP/IP Internet typically admits these nodes without restraint, and any node that can access the network can join. Networks can refine admission policies, however, to restrict membership to interest groups. Likewise, the operation domain can be a controlled environment such as an intranet rather than the public Internet.

P2P networks can use protocols that make it easier for individual nodes to participate and share information, but the trade-off can be decreased security or quality of service. For example, using a lightweight alternative such as the user datagram protocol may sacrifice the reliability that a protocol such as TCP offers.

> **The P2P paradigm can be extended and enhanced to foster productivity in the workplace and support community activities.**

Individual nodes connecting to the network can access a real-time index of other active nodes and of the files they share. In turn, as soon as they connect, these nodes become part of the index themselves, with the files they choose to share automatically added to the index. Because the index provides addresses for resources available at any given time, a member node can simply initiate a direct connection with any connected member node that currently holds the requested information. Protocols for initiating such connections are part of P2P networks.

In a file-sharing network like Napster, having a central node host the index adds some degree of centralization. Individual nodes access this central node to look up the peer node that hosts the content and then engage in a P2P interaction to acquire the content directly from the peer node. When a network operates without a central node, each node can have a partial index of member nodes it knows. A lookup for content can start with this index and propagate to directories found at other nodes. Such a system is less efficient in search and retrieval, but it is less vulnerable to a central node's failure, and it will not load up any single node excessively. In either case, a member node sees an index of currently available information, at currently active nodes, and doesn't need to sort through dead or outdated links.

Most of these P2P network features translate into a rather unstructured, almost primitive information-sharing network that quickly builds itself into a comprehensive information collection. This structure also allows effective sharing of information by large communities at low cost.

P2P's downside stems from two factors:

- the possibility that the listed information will be cluttered with significant noise, and
- inefficient directory lookup operations in the case of arbitrary organization and distribution of content.

Within a controlled environment that can correct for some of these factors, however, a P2P network can be an effective information-sharing and community-building tool.

The P2P networks widely known today engage primarily in sharing music or video clips or gaming software—activities not directly perceived as positively contributing to electronic commerce. However, the P2P paradigm can be extended and enhanced to foster productivity in the workplace and support community activities. Most of the applications we envisage differ from what P2P networks have been used for in the past.

UNLEASHING THE DECENTRALIZED NETWORK

Several advantages make P2P networks attractive.

Enhanced load balancing

A P2P environment can use a wide range of policies for distributing information. The more refined load-balancing techniques monitor traffic and the demand profile for particular information items, then redistribute content to ease the load on individual nodes and possibly locate content closer to points of high demand. Proactive load-balancing schemes can effectively ensure, for example, that a user does not need to connect to a site on the other side of the globe when the same file resides on a node in the same building.

These load-balancing techniques represent a relatively less sophisticated and computationally expensive solution than distributed caching.[1] P2P allows effective distributed storage of a wide variety of content that can find caching techniques too expensive. Content relocation is especially effective if the network automates the process by dynamically monitoring usage patterns and replicating content to balance the load optimally. The "Two Potential P2P Applications" sidebar, for example, describes a potential automated load-balancing environment, while current initiatives such as Freenet already aim to deploy automated content relocation and replication policies. In addition to achieving load balancing, automated content relocation could also make searches more effective.

Dynamic information repositories

Any P2P network user can scan the active nodes for desired information, then download it directly from the node. Users who download information can make it available for sharing from their own nodes to others. Thus, any information in high demand can rapidly spread to many nodes. As the network grows, the amount and scope of content available for sharing grow as well. In an open-access scheme, a popular network could quickly accumulate a comprehensive collection of content. A user searching for even the most rare MP3 file of an old song could eventually find it at some node. This aspect of P2P interests researchers because it offers content-based addressing, leading to the notion of more effective—as opposed to more efficient—routing.

Redundancy and fault tolerance

Replicating information at multiple nodes provides a high degree of redundancy, which in turn leads to a high degree of availability and makes it possible to serve more users. Further, redundancy implies that no single node can render the network dysfunctional if it fails. Decentralization thus enhances fault tolerance and security. These advantages, similar to those of the original Internet design, reflect Napster creator Shawn Fanning's statement that the file-sharing network "is a throwback to the original structure of the Internet."[2]

Content-based addressing

On the Web, URLs address resources and may not always directly relate to their specific content. In P2P, the exact address of the node that stores a particular item remains transparent to the user. The user queries the network for the content and the P2P software translates the request into the specific nodes that hold the content in storage. This procedure can lead to a grouping of addresses based on the content the respective nodes store. Addressing reaches a higher level in the semantic hierarchy because users specify a content identifier but not a physical location. The identifier corresponds to a P2P collection of nodes that store this type of content. Thus, segregation of content into specialized groupings distributed over P2P networks can create a more refined information repository and more uniform resource identifiers.

Improved searches

Fanning says that his original motivation for developing Napster came from problems his friend encountered with Web search engines.[2] These search engines rely on Web crawlers that scour the Internet for content and store the information in massive, searchable databases. Such indexing only includes content from publicly operating servers, and the databases do not undergo immediate updates when any of those servers or links go down.

In P2P, by contrast, the information available at any user node is indexed, and remains indexed, only when the user is online. Consequently, the index always synchronizes with current status. Unlike search engines, P2P indexing begins when the user logs into the P2P network and sends a content-index message. A P2P network does not rely on search engine robots to revisit its links and update the information. This dynamic indexing of content and the ability to easily search that content distinguish P2P from other applications, such as newsgroups, which also allow information sharing among communities.

In a P2P network, search efficiency depends on how well organized the network is, its policies for classifying content and building directories, and the search software itself. To develop a business P2P network, developers could use representations such as XML to standardize descriptors and organize the content. Such

Two Potential P2P Applications

Widely publicized for its music-sharing abilities, P2P technology could easily handle other, more challenging tasks.

P2P Multicasting

Consider a multimedia streaming content provider that uses multicast to distribute content to its customers. Multicast sends out only one copy of data from the source, then replicates it as far downstream as necessary, taking advantage of the overlap in distribution paths. Each node that replicates effectively acts as a secondary source.

Local ISPs could specialize in customized multimedia content and push it to customer communities by multicast. The local community that prefers a particular type of content, say sports clips, subscribes to the local ISP-content provider that specializes in that content. The ISP, in turn, has a P2P network of retail nodes that it populates with content based on traffic and demand patterns. The customer obtains content through a client-server interaction with the central directory at the provider that identifies the optimal download location and then downloads content from this location.

Although the customer-provider and customer-retail-node interactions are not peer-to-peer, the customer benefits by downloading content from a highly available, redundant network of nodes in geographic proximity. Further, the customer avoids dealing with congestion patterns in the backbone and with interdomain quality of service (QoS) considerations.

P2P Computing and Active Networks

Active network developers envisage transforming IP packets into encapsulated fragments of executable code that traverse the network and execute in limited environments at intermediate nodes. In essence, this concept significantly advances the level of current networks' sophistication.

Peer-to-peer networking, on the other hand, was devised as a lightweight, primitive, networking concept that achieves adequate results at minimum cost without sophisticated network protocols. At first sight, then, it seems ironic to link the two. However, we could easily think of active capsules that carry P2P queries to locate particular content while also determining its metrics, such as time of modification, size, or QoS specs.

Further, active capsules could carry code for determining traffic and usage patterns for individual products at each node and dynamically make decisions on redistributing content across the P2P network. Such a combination of P2P and active networks would be best suited for business communities. These communities impose some form of central control on nodes that can participate and enforce guidelines on the type of mobile code the system executes.

Because P2P queries can be more lightweight than distributed database queries, bypassing interoperability and legacy system issues, the mobile code would pose a minimal computational burden and impose minimal network overhead. Such a combination could lead to highly efficient self-sustaining and self-maintaining P2P networks within organizations.

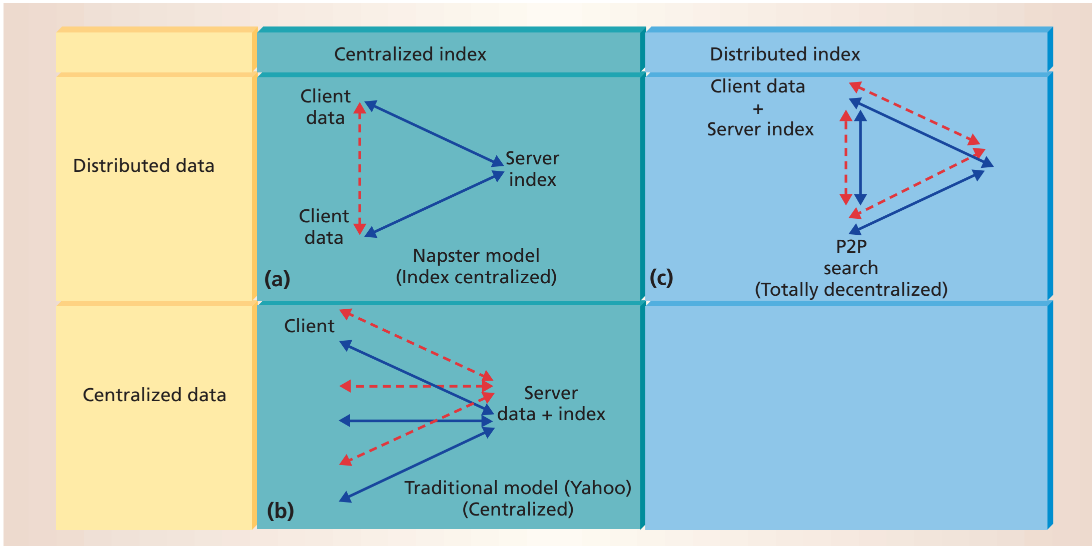


Figure 1. Search model comparison. The dashed lines indicate data transportation, while the solid lines indicate search transportation. (a) Napster offers a centralized index and distributed data, (b) the typical Web search engine centralizes the index at its own server and the data at each content server, and (c) a fully distributed P2P application distributes data and searches.

standardization would facilitate the automation of content replication and relocation, allowing more effective searches and providing dynamic responses to changing demand profiles. Figure 1 compares the data- and search-transportation profiles of Napster, a typical Web search engine, and a fully distributed P2P application.

BACK TO CHAOS

There is, of course, a flip side to the story. The emphasis on decentralization and user autonomy that gives P2P its advantages also stirs up potential problems. However, if we carefully direct our efforts toward specific purposes in a controlled manner, we can indeed bottle this genie.

Spurious content, poor connections

P2P's extreme flexibility could be its own undoing. Because the network admits individual nodes without restriction, the quality of their links and the capacity of their servers can vary widely. A user who shares files through a dial-up connection may have a low-speed, failure-prone connection. Another user may have a low-end PC that cannot support a high traffic volume. There is thus no way to verify or standardize information or to ensure its authenticity.

Consequently, a song title listed on multiple nodes can have such varying quality that some versions completely lack authenticity. Although P2P implements content-based addressing, specific content listings will likely include spurious items because P2P's decentralized nature implies that a central administrator cannot remove objectionable or invalid content. The advantages of P2P searches may well be undermined if the links they point to vary widely in content and connection quality and they aren't associated with any certification or standards. The search will be effective, but could be inefficient.

Various measures can address these problems. In a public P2P network such as Napster, information on connection speeds and node characteristics can classify content. A network can apply user feedback to build contributor profiles, similar to the scores eBay provides for auctioneers. Napster has policies that allow it to ban users, while some P2P software groups monitor users who share objectionable material and warn the community about them.

More significantly, a private P2P network in which users exchange decision-making information can control membership, implement authentication schemes, and adopt standardized schemes to describe and classify content. These networks can also exert control over the infrastructure they use and the reliability of connections. Such standardization can render searches efficient as well as effective.

Traffic redistribution

P2P networks can redistribute traffic from heavy-load sites to other nodes that replicate the information. For such shifting to reduce traffic, the source must be located close to the destination. To find such nodes, we need algorithms that determine which peer has the required information and is closest to the destination. However, uncovering this information requires so much computing power that a perfect solution may be infeasible.

As demand varies, nodes must make decisions about which new content needs to replicate and what the content replaces. To deal with such issues, cache-replacement schemes use complex algorithms to repopulate faster, temporary storage with frequently demanded content to speed access to slower, more permanent storage. Distributing replication makes the problems more complex and requires more computing power.

These concerns do not mean that available computational resources cannot fulfill P2P's promise. Researchers are already using replication algorithms to develop P2P protocols that provide improved load balancing. Although not necessarily a perfect solution, P2P load-balancing schemes are expected to be adequate for the constituencies they serve and the type of information they share.

Free riders

P2P computing's decentralized nature makes it ideal for economic environments that foster knowledge transfer. Such communities are vulnerable to the classic free-rider problem:[3] Any node in a P2P network can steal quality links or information from other nodes by replicating the information and offering it as a bundled-product component. Often, we have no way of differentiating the original sources of content; even if we did, customers may not care about the source as long as they receive quality data.

P2P computing's decentralized nature, ease and speed of growth, and open-access policies may make regulation of content propagation an insurmountable challenge. Software designers must think of ways to regulate the growth of P2P networks while ensuring that content providers in the P2P network receive compensation for their efforts.[4] Cohesive social groups that impose norms on public behavior can ensure that users create and distribute content equitably. However, the key problem is restricting participation. If we don't exclude people who might want a free ride on the network, content may be undersupplied. To ensure a smoothly running P2P network, we propose implementing cost-sharing mechanisms. For example, a user who does not contribute content can at least lend CPU unit cycles. Alternatively, an internal currency system can effectively privatize a public good by compensating contributors with credit toward network content and services.

The wild, wild Web

The legal battles over Napster and Swapoo.com have shown that letting any user freely share any type of information raises issues of copyright infringement, intellectual piracy, and the potential spread of undesirable content such as child pornography. Yet P2P proponents advocate imposing no controls on what type of information users share.

P2P computing's decentralized nature, ease and speed of growth, and open-access policies may make regulation of content propagation an insurmountable challenge.

This lax approach has led P2P into the quagmire of litigation and regulatory headaches. Currently, P2P users can digitize copyrighted or confidential material and disseminate it worldwide. Malicious users can distribute undesirable material in fly-by-night operations that, by the time tracing activity detects them, may no longer be part of the network.

Short of imposing judicial or regulatory actions that shut down the entire network, public communities can solve these problems only through self-regulation. Indeed, several P2P communities are already taking steps in this direction. They reserve the right to banish or blacklist users known to distribute undesirable material and also enforce guidelines on what type of content members can share.

While the notion of self-regulated communities may seem idealistic, given that a community's value to its members depends on its nature and reach, powerful incentives exist for excising users or content that stray from the community's mainstream. Rogue or fringe communities can still form, of course. But the Internet has always hosted such communities, from the early days of Internet Relay Chat to more recent newsgroups. In each case, as the technology and community matured, their beneficial effects have dominated harmful ones, minimizing the impact of rogue communities.

Beyond social regulation, other factors contribute to network providers denying service to such elements, such as aggressive litigation by interest groups and affected parties. Such groups usually target the network service provider that hosts the site, rather than the content provider itself—who may sometimes be more difficult to trace. Providers have typically responded by quickly shutting down any reported or suspected content or activity that's likely to be objectionable or to invite legal or government action. Further, governments have become increasingly active about intervening in such cases.

Whether the freedom to share information can be limited is the subject of a continuing debate that transcends Internet communities. An immature and highly publicized technology, P2P forms the basis for many chaotic and lawless communities. As they evolve, we can expect P2P communities to become more stable and have fewer rogue members.

Cracker heaven

Nevertheless, P2P introduces new pitfalls for the unwary and new opportunities for the unscrupulous. P2P networks require that users share their files publicly. Given the average user's limited awareness of security issues and the lack of stringent security fea-

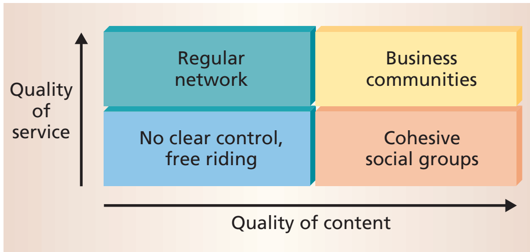


Figure 2. Peer-to-peer service quality matrix. The complex interplay between economic, technical, and social influences shapes the nature of P2P networks.

tures in PC operating systems and many P2P protocols, crackers—malicious users who pirate information from systems, infect them with harmful computer viruses, or both—could exploit this vulnerability.

In addition to the threat to individual users, we must consider the more disturbing possibility that user sites could become launch pads for distributed denial-of-service attacks. Many P2P protocols sacrifice security features to prune overhead. Using such architectures in business communities that share critical information could lead to serious security vulnerabilities.

As new applications expose potential security problems, software vendors have quickly devised measures to address them. Consider, for example, the security issues that cable modems and DSL raise and the resulting widespread availability of firewall solutions. The same pattern may benefit P2P users. Business P2P platforms, for example, can use the secure sockets layer protocol or other solutions to build more secure architectures.

Picking lemons

P2P networks will lose value if they have poor QoS and quality of content. Unfortunately, P2P computing seems to invite content that scores low in both areas. Thus, users may be faced with sorting through lots of worthless content before they find something valuable.

Many P2P user interfaces address QoS concerns by offering information on the type and speed of connections that serve particular content. User-feedback-based scores can provide information for enhancing connection reliability and content quality. We suggest studying P2P networks from an integrated perspective that encompasses assurance of delivery, quality of content, and QoS.

The user derives value from a combination of all three factors. A highly reliable Web server has no value if the content the user requests cannot be found there. A high assurance of content delivery is useless if the user can't get a good connection to the source node.

These issues become more tractable in a controlled membership environment, such as a business community or a self-enforced, restricted-membership social community like a neighborhood P2P network. In such an environment, users can impose standards on quality of content and service. Active participation from the user community assures high-quality service and content delivery. On the other hand, open membership in public communities contributes to the content's comprehensiveness, thereby yielding an almost deterministic assurance of delivery.

P2P COMMUNITIES

To use this emerging technology effectively, we must understand the complex interplay between economic, technical, and social influences that shape P2P networks, as Figure 2 shows. Their decentralized nature implies that the technology's performance and overall productivity depend on the web of social interactions that characterize P2P communities. The P2P information-sharing paradigm is attractive to social groups and business communities. The nature of the P2P community influences the economic transactions of the group, since group behavior depends not only on the behavior of other group members, but also on the context of interaction within the members of the group.[5] For example, a social group displays more altruistic behavior and relies on social norms for enforcement, whereas the regulation of a business community would be based on more rational rules governing transactions.

P2P communities can be social or business communities. Implementing a quality assurance mechanism can take place only within the context of the community. For a P2P group to be cohesive, the group's users must share similar ideas on how they want to interact with each other. Viewed from the community perspective, P2P network participants derive value from both the interaction with a peer group that shares a matching interest and from the products they share.[6] This relationship in turn influences the flow of P2P information, as users determine how to cooperatively share resources that are publicly available to everyone. Social norms influence individual behavior in a group to the extent that the group's members value conformance to those norms—a relationship that P2P networks can use to enforce some sort of self-regulation.

Social communities

P2P social communities form by themselves. Likewise, grouping customers according to demand profile occurs via self-selection. The lack of centralized control means the products the community's members exchange can be of inconsistent or even inferior quality. Further, participant nodes and their network connections can exhibit varying levels of capability.

Despite these shortcomings, the technology's advantages dictate that market segments that can tolerate deterioration in quality will form P2P communities. Napster offers a good example. Whereas the MP3 format may not offer CD-quality music, Napster users gladly listen to MP3s for free. Further, the musical selections that the Napster community shares rarely demand the subtle improvement in audio quality that CDs offer.

Indeed, a P2P social community can share many socially beneficial digital goods that do not require rigorous quality control. For example, K-12 educational communities could use P2P networking to exchange reference information and instructional material. Each residential community could form P2P networks to aid information sharing. Likewise, informal P2P gatherings can provide opportunities for exchanging recipes, sharing feedback on various products and services, or providing other information helpful to consumers.

Research in experimental economics can help analyze P2P computing groups in terms of the self-selection and social-enforcement mechanisms that can enforce quality of delivery. We can follow the precedent set by work on reinforcement and learning mechanisms[7] in determining how a P2P network propagates itself.

Business communities

The advantages of using P2P computing for business include the ease of implementing, joining, and using such a network. Further, P2P networking offers savings compared with distributed databases and avoids the headaches caused by dealing with interoperability and legacy-system issues. Organizational control will remedy P2P's inherent difficulty in providing certifiable quality of content and predictable quality of delivery.

Although P2P technology is less efficient than a distributed database or real-time decision support system, its specific application is for sharing content that does not require real-time or mission-critical handling. The controlled membership domain implies that the P2P network remains internal and that arbitrary anonymous nodes do not participate.

Incentive is the main organizational problem that limits knowledge sharing between people. For example, a network administrator in Corporation XYZ may be unwilling to take the time to post tips on network maintenance that other users could then access through the collaborative file-sharing system.

To overcome this behavior, organizations could introduce payment mechanisms to implement P2P information-sharing systems that enhance decision-support processes. Perhaps the network could provide a payment and settlement service to ensure that content creators receive compensation.

Say, for example, user A needs information on sales data and is willing to trade her expertise in accounting to get it. The sales department may be willing to give her the information she needs, but they have no use for accounting data. To resolve such a situation, the company could set up a payment service that lets people trade their knowledge for a fiat currency. User A gets points from an employee who needs accounting advice, then redeems those points with the sales department, which could in turn use the points it earns to buy lunch at XYZ's cafeteria.

The advantages of using P2P computing for business include the ease of implementing, joining, and using such a network.

Businesses can view information transfers as investments and measure the value of information in terms of the indirect effects it might have upon the recipient's efficiency. More specifically, businesses can use monitoring mechanisms in conjunction with the payment and settlement system to trace the value of information to the participants. A decision maker in the organization can act as a trusted third party responsible for assigning values to the knowledge being shared. Indeed, at least one commercial initiative, Mojo Nation (http://www.mojonation.net), promotes the use of a payment mechanism for solving IT-related managerial problems over a P2P knowledge-sharing scenario.

The key to realizing P2P's promise lies in the availability of enhanced services such as structured ways for classifying and listing shared information, verification and certification of information, effective content distribution schemes, and security features. Customized P2P networks may best serve the needs of large organizations with specific requirements, whereas smaller organizations and social communities might benefit from using software that third-party vendors provide.

P2P offers a superior alternative to the client-server model for at least a subset of that technology's applications, as evidenced by the phenomenal success of P2P music-sharing networks versus centralized MP3 servers. Even so, it is unlikely that P2P networks will replace the client-server paradigm. Rather, P2P will offer exciting new possibilities in distributed information processing. ✷

Acknowledgments

The authors thank Intel and IBM for their support.

References

1. A. Gupta, D.O. Stahl, and A.B. Whinston, "The Economics of Network Management," *Comm. ACM*, Sept. 1999, pp. 57-63.

2. S. Fanning, Testimony before Senate Judiciary Committee, Provo, Utah, 9 Sept. 2000; http://dl.napster.com/ Dec_Fan.pdf (current 21 May 2001).
3. S. Ba, J. Stallaert, and A.B. Whinston, "Optimal Investment in Knowledge within a Firm Using a Market Mechanism," to be published in *Management Science*.
4. T. Groves, "Incentives in Teams," *Econometrica*, vol. 41, 1973, pp. 617-663.
5. C.F. Manski, "Economic Analysis of Social Interactions," *J. Economic Perspectives*, vol. 14, no. 3, 2000, pp. 115-136.
6. E. Ostrom, "Collective Action and the Evolution of Social Norms," *J. Economic Perspectives*, vol. 14, no. 3, 2000, pp. 137-158.
7. D. Stahl, "Rule Learning in Symmetric Normal-Form Games: Theory and Evidence," *Games and Economic Behavior,* vol. 32, 2000, pp. 105-138.

***Manoj Parameswaran** is an assistant professor of decision and information technologies at the Smith School of Business, University of Maryland. He holds a joint appointment with the Institute for Systems Research, University of Maryland, and is a research associate at the Center for Research in Electronic Commerce at the University of Texas at Austin. He received a PhD in management information systems from the University of Texas at Austin. Contact him at pcm@rhsmith.umd.edu.*

***Anjana Susarla** is a doctoral student in the Department of Management Science and Information Systems and a research associate at the Center for Research in Electronic Commerce, both at the University of Texas at Austin. She received an MBA from the Indian Institute of Management, Calcutta, India. Contact her at anjana@mail.utexas.edu.*

***Andrew B. Whinston** is the Hugh Roy Cullen Professor of Information Systems, Economics, Computer Science, and Library and Information Sciences and is the director of the Center for Research in Electronic Commerce, both at the University of Texas at Austin. He received a PhD in economics from Carnegie Mellon University. Contact him at abw@uts.cc.utexas.edu.*