

**OF TRANS-DISCIPLINARY CONCERN:  
A MULTI-DIMENSIONAL  
CLASSIFICATION OF THE  
PEER-TO-PEER FILESHARING  
RESEARCH LITERATURE**

**Jerald Hughes**

#CIS-2003-05

**November 2003**

**Jerald Hughes**

Doctoral Student, Specialization in Information Systems  
PhD in Business  
at Zicklin School of Business, Baruch College  
and the Graduate Center of the  
City University of New York  
Jerald\_Hughes@baruch.cuny.edu

©2003

---

The CIS Working Paper Series provides for early dissemination of research produced by members and affiliates of the Computer Information Systems program at the Zicklin School of Business, Baruch College, City University of New York and the Information Systems program of study of the PhD in Business program of the City University of New York.

Please address all inquiries to:

CIS Working Paper Series  
Zicklin School of Business  
Baruch College, Box B-11-220  
17 Lexington Avenue  
New York NY 10010

**OF TRANS-DISCIPLINARY CONCERN:  
A MULTI-DIMENSIONAL CLASSIFICATION OF THE  
PEER-TO-PEER FILESHARING RESEARCH LITERATURE**

ABSTRACT

*Since the advent of Napster and the Peer-to-Peer (P2P) filesharing applications which followed it, distributed storage and sharing of content has been a topic of increasing interest in many different disciplines. This paper presents a review of the literature on P2P filesharing systems. The research literature on such systems crosses the boundaries of several disciplines, so a multidimensional classification scheme encompassing all the relevant fields addressed in the existing literature is employed. A database design reflecting this classification system is discussed, and has been implemented. The potential uses of the multidimensional classification are explored. Patterns of research made visible by queries on the database are discussed. Findings of the various research fields are summarized. The contributions of this paper are (1) identification and implementation of frameworks for exploring research specifically on P2P filesharing systems, (2) the creation and implementation of a meta-framework approach for categorizing IS literature, and (3) a summary of the current state of research on P2P filesharing systems.*

# OF TRANS-DISCIPLINARY CONCERN: A MULTI-DIMENSIONAL CLASSIFICATION OF THE PEER-TO-PEER FILESHARING RESEARCH LITERATURE

## Table of Contents

### 1. Introduction

1.1 <a href="#">Peer-to-Peer Filesharing</a> .....	1
1.2 <a href="#">Motivation and Selection</a> .....	3
1.3 <a href="#">IS Classification Schemes</a> .....	4
1.4 <a href="#">A Trans-Disciplinary Literature</a> .....	7

### 2. P2P Citations Database

2.1 <a href="#">Structure</a> .....	11
2.2 <a href="#">Classification Frameworks</a> .....	13
2.2.1 <a href="#">Inside or Outside the Firm?</a> .....	16
2.2.2 <a href="#">Methods and Approaches</a> .....	20
2.2.3 <a href="#">Perspectives</a> .....	22
2.2.3.1 <a href="#">Digital Rights Management</a> .....	22
2.2.3.2 <a href="#">Computer Science and P2P Performance</a> .....	25
2.2.3.3 <a href="#">Human Computer Interaction, Ethics</a> .....	31
2.2.3.4 <a href="#">Law</a> .....	33
2.2.3.5 <a href="#">Behavior</a> .....	34
2.2.3.6 <a href="#">Economics</a> .....	37
2.2.3.7 <a href="#">Strategy</a> .....	38
2.2.3.8 <a href="#">Culture, Marketing</a> .....	39

2.2.4 <a href="#">The IT Artifact</a> .....	41
2.3 <a href="#">Locating the Research</a> .....	44
2.4 <a href="#">Filesharing Ratings</a> .....	46
3. <a href="#">Conclusion</a> .....	49
<a href="#">References</a> .....	52
<a href="#">Appendix</a> - Database Construction Conventions.....	60

## List of Tables and Figures

### TABLES

Table 1 - 1993 Classification Systems for IS Research.....	5
Table 2 - Performance Issues.....	29
Table 3 - Filesharing Ratings by Discipline.....	48
Table 4 - Industry Strategies Versus P2P Filesharing.....	50

### FIGURES

Figure 1 - Positioning the Literature.....	3
Figure 2 - Relationships Structure of the Database.....	11
Figure 3 - P2P Filesharing Studies.....	13
Figure 4 - Frameworks with Associated Types.....	15
Figure 5 - Inside or Outside the Firm Counts.....	17
Figure 6 - Filesharing Applications for Digital Media.....	18
Figure 7 - Methodologies and Research Approaches.....	21
Figure 8 - The DRM Nexus.....	23
Figure 9 - Paper Counts by Perspective.....	27

<b>Figure 10 - Free Riding Behavior.....</b>	<b>35</b>
<b>Figure 11 - Piracy Behavior.....</b>	<b>36</b>
<b>Figure 12 - IT Artifact Framework Distribution.....</b>	<b>43</b>
<b>Figure 13 - University Departments of the P2P Authors.....</b>	<b>44</b>
<b>Figure 14 - Research Disciplines of the Publishers.....</b>	<b>45</b>

## 1.1 Introduction - Peer-to-Peer Filesharing

Since the advent of Napster and its successors, Peer-to-Peer (P2P) filesharing systems have become a topic of great interest in many research fields. The literature which is the topic of this paper examines P2P systems from several perspectives, but falls far short of exhausting all of the possible and proposed applications of P2P technology.

A *Peer-to-Peer* system is one in which the nodes of the network communicate as equals, with no centralized governing node as a go-between (Whinston et al, 2001). Each participant in the network can behave either as a client, receiving files, or as a server, sending information, or both, and can establish these relationships with any other node in the system. Peer-to-Peer architecture is basic to the structure of the Internet, as implemented in the TCP/IP protocols (Leiner et al, 1997). The P2P systems we will look at should be considered overlays upon that basic Internet infrastructure. Within the notion of P2P as overlay networks, there are several basic topics of interest (Damiani et al, 2002):

Filesharing - the distributed storage and sharing of digital content, including but not limited to music, movies, photos, software, and books. These are the systems such as the original Napster, Kazaa, Grokster, and so forth. There is also work in this area being done to make P2P systems ready for other uses-- P2P systems for data management, marketing and distribution, and content access.

Distributed Processing - P2P overlay networks can organize the computational resources of thousands or even millions of computers, to achieve supercomputer levels of performance. An early example of this is SETI@Home, which uses the

donated spare CPU cycles of volunteer participants' computers to analyze radio signals from outer space, in search of patterns which may indicate extraterrestrial intelligences.

One-to-One Communications - Instant Messaging systems, such as those implemented by AOL and others.

Media Streaming - delivering high-bandwidth content through the cooperative actions of all the machines in the network, instead of through dedicated servers.

Ad Hoc Mobile Networks - mobile computing platforms can participate in ad hoc networks via wireless communications protocols, including IEEE 802.11 (Conti & Giordano, 2001).

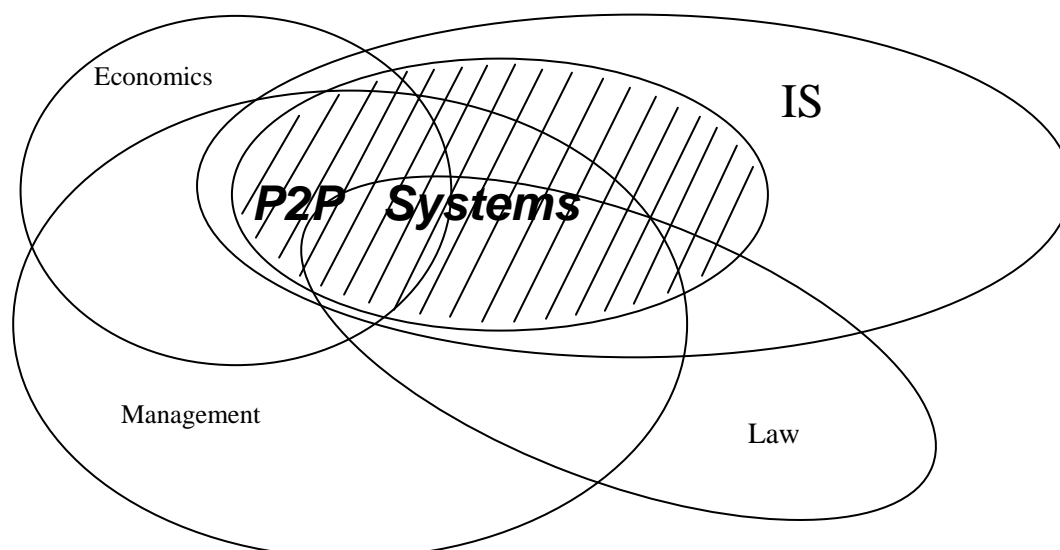
The notion of P2P overlay networks in general is attractive from a performance-oriented point of view, since they solve so many problems. A distributed processing network, for example, is essentially a massively parallel computing architecture which leverages the otherwise idle CPU cycles of participating platforms (Casanova, 2002, Foster 2003). Media-streaming P2P systems can leverage the power of otherwise unused bandwidth to achieve cost-effective, fault-tolerant content delivery (Hefeeda et al, 2002 and 2003; Habib and Chuang, 2004; Habib et al, 2004). Web content servers which are presently vulnerable to Denial-of-Service attacks, because data is served from one or a few static, known locations, could protect themselves from such disruptions of service by storing the content in a distributed manner, using P2P technologies. All of the areas in the list above appear to be of great interest to Computer Science (CS) researchers developing new P2P applications, but this review will be confined to articles related to P2P filesharing systems only, systems in which the users freely exchange

content directly between themselves. The most well-known of such systems to date are music filesharing networks such as the original Napster or Kazaa, but there are dozens more already operating (BitTorrent, WinMX, Poison, e.g.), and many more in development. These systems are now being used to exchange not only music, but also movies, television shows, books, software, pornography, and virtually anything else that can be digitized. Articles which discuss topics directly relevant to filesharing activity, such as technical innovations intended to improve the efficiency of Gnutella-type query resolution for example, are also included.

## 1.2 Introduction - Motivation and Selection

The papers included in this review were selected for the purpose of supporting an IS-oriented research agenda, one which concentrates on the links between the characteristics of IT, its developers and users, and business outcomes. In examining just the work on Napster alone, one immediately finds that a proper understanding of the forces at work requires contributions from separate research disciplines. The perspective of P2P systems presented herein thus lies at the intersection of several distinct bodies of literature (Figure 1):

**Figure 1 - Positioning the Literature**



[Note: Figure 1 is for conceptual purposes only, and is not intended to represent relative size, or proportion of overlap between fields. For the purposes of clarity, two other disciplines which have also made contributions to our understanding of P2P systems, Marketing and Sociology, are omitted. A representation which was both complete and correct in all of its proportions could require a display in 5 spatial dimensions.] Common to all the shaded portions of the illustration is the IT artifact: P2P filesharing systems. In the literature examined here, we see links between IS and Economics, IS and Law, IS and Management, IS and Marketing, IS and Sociology, and any combination of the above.

This sector of IS research has several attractions. It is an area of recent and swift innovation, with novel applications of IT which have important consequences for business, both pro and con. The literature on P2P systems which addresses filesharing demarcates a well-defined subset of the possible applications of P2P overlay technologies in general, and includes work on mature, fully developed systems in common use and therefore available for study; this is as opposed for example to P2P systems used for media streaming, which are still in development, or those used for grid computing, which is an important area but not one of primary interest for the author, whose background includes experience in multimedia and music, including both performance and digital audio recording and production. It is hoped that this firsthand engagement with the issues involved will lend itself to the realization of useful insights.

### **1.3 Introduction - IS Classification Schemes**

Barki et al published a comprehensive keyword schema in *Management Information Systems Quarterly* in 1993. This system identified nine broad categories, with dozens of sub-categories in a four-tiered hierarchical system. At about the same time (1993), Swanson and

Ramiller published a review of submissions to *Information Systems Research* which used a bottom-up approach of increasing abstraction to identify nine overall themes.

**Table 1 - 1993 Classification Systems for IS Research**

<b>Barki, Rivard &amp; Talbot - MISQ</b>	<b>Swanson &amp; Ramiller - ISR</b>
Reference Disciplines	Information and Interface
External Environment	Computer Supported Cooperative Work
Information Technology	Introduction and Impact
Organizational Environment	Systems Projects
IS Management	Users
IS Development and Operations	Evaluation and Control
IS Usage	Economics and Strategy
Information Systems	Decision Support and Knowledge-Based Systems
IS Education and Research	IS Research

Of the categories identified by Barki and his colleagues, P2P filesharing studies are well-represented in all but two, Organizational Environment and IS Education and Research. Since P2P systems have not yet been adopted for the purposes of realizing institutional goals, there has so far been little or no opportunity for studying, say, 'Information Flows' (keyword code DD07

under MISQ's scheme) as implemented by an intra-organizational P2P system. However, if the P2P development research observed is any indication, then such opportunities may arise in the future.

The P2P research is represented in all of Swanson and Ramiller's categories except Decision Support and Knowledge-Based Systems, and IS Research. As of this writing, working P2P systems to support Computer Supported Cooperative Work have not been implemented, but design issues and possible architectures are being explored. For example, Park & Hwang (2003) have designed a P2P architecture which provides for access control for collaborative systems; Edwards et al (2002) have designed a P2P system, SpeakEasy, which allows the formation of ad hoc workgroups using P2P communications. P2P filesharing studies have been done on Human-Computer Interaction under 'Information and Interface', on data modeling (replication strategies) under 'Systems Projects', on IS Security and IS Ethics under 'Evaluation and Control', on User Perceptions and Attitudes under 'Users', on IS Economics under 'Economics and Strategy', and on P2P Adoption under 'Introduction and Impact'. This paper would constitute a first entry in the MISQ system under 'IS Research Frameworks' (code IB02) in the IS Education and Research area, or in the ISR system as "generalization of results" under 'IS Research'.

On the other hand, it is clear from the context in which these classification systems are offered that their authors likely had particular kinds of business-oriented information systems in mind to which their categories would mostly apply, systems which have some characteristics in common with P2P filesharing systems, but which are quite different in other respects. Since 1993, new general areas of interest to IS researchers have sprung up, such as E-commerce, which introduce problems (such as customer trust and website effectiveness) captured only imperfectly by the MISQ and ISR classifications. P2P filesharing systems complicate the picture even more,

because they not only raise issues associated generally with Internet applications, but also create new problems, problems which are specific to systems which have arisen outside of any particular institutional context.

#### **1.4 Introduction - A Trans-Disciplinary Literature**

P2P filesharing systems are thus far unique as a concern in Information Systems (IS) research, in that they are fully functioning, immensely popular, widely used and applied networks which have enormous potential consequences for businesses both from the inside, as deliberately implemented systems to achieve institutional goals, and from the outside, as challenges to existing business models, under no particular control. For most of the history of computing, the only entities capable of launching powerful new computer systems were institutions with concentrated resources, such as government agencies, businesses, and universities. This purely practical limitation is demonstrably no longer in place; the flexibility and power of computers available to consumers, in combination with the high degree of connectivity enabled by Internet technologies, has conferred upon Internet users in general the power to implement self-organizing cooperative systems. Taking advantage as they do of users' already existing hardware and software, P2P filesharing systems can be launched simply by distributing an application, such as Kazaa. Though it has been used by millions, Kazaa is not necessarily a system completely governed by an institution which owns and operates it, like AOL. In fact, depending on the specific details of how they are built, P2P filesharing systems may not be governable at all; once the application is released onto the Internet, it may exist independently of its originator and of any attempts to control it.

The result is that it is not really possible to speak authoritatively about P2P filesharing systems from entirely within the confines of a single discipline. There is a swiftly growing body of literature in Law journals, for example, which addresses itself to the legal issues of copyright raised by Napster and others. The conclusions drawn by U.S. courts, and by lawyers examining the issues, necessarily turn upon rather fine points of computer technology of the sort which is developed and discussed in the CS literature. Beginning with the first challenges to Napster in the courts, a sort of legal/technological arms race has arisen, in which business organizations such as the Recording Industry Association of America (RIAA) resort to legal means to shut down the Napster music filesharing network, whereupon software developers rework the architecture of P2P systems to avoid precisely the legal points cited in the decision--any P2P system which is an implementation of the Gnutella filesharing architecture, such as Kazaa, has no central index, as Napster did. The RIAA then sues users for copyright infringement, and P2P developers respond by creating anonymous systems or by distributing single files among thousands of users (the BitTorrent system leverages user bandwidth by allowing one to download a particular video file, for example, from many different users--there is no particular single source for the file). At least one system explicitly purports to confer 'plausible deniability' among its users:

*"The GA concept works by allowing users to instigate searches on behalf of a group of users so that no individual can be linked to the results of any of those searches. In other words, every user can plausibly deny having been the instigator of any particular query." - Krishnan & Uhlmann, 2004*

The RIAA's concern is control of copyrights; in order to understand their situation, one must refer not only to the legal concepts of copyright and fair use, but also to Digital Rights Management, which itself brings in issues of Computer Science (is there even such a thing as

secure control of content, and if so, how can it be implemented), Economics (incentives, profit-maximization), and Strategy (how can profits be maximized under various conditions of content piracy). If we look at the users of P2P systems, then questions immediately arise of Ethics (how is the decision to acquire content under copyright made?), Culture (what are the values and belief which characterize users of P2P systems?), and user Behavior in general (do they share files? how many? of what kind? how long do they stay connected to the network?, etc.). The answers to questions in any one area affect and are affected by the questions asked in other areas. For this reason, we can characterize the study of P2P filesharing systems as *trans-disciplinary*: in order to effectively study them, we need to have a grasp of how different research disciplines can contribute both singly and in combination to our understanding. Swanson and Ramiller (1993) did explore models of research themes which pointed out important links between different areas, but finally chose to classify each of the papers they looked at into only one category, the category considered the central theme of that particular paper.

This review attempts to use database technology to implement a different approach, in which no final judgment of a single category is necessarily assigned to a particular paper. Instead, all of the topical categories into which a paper fits are identified and implemented as many-to-many relationships. For example, Chen and Png's (2003) paper "Information Goods Pricing and Copyright Enforcement Welfare Analysis" brings together both Ethics and Economics to study the impacts of P2P filesharing systems which exist 'Outside the Firm', and thus is keyed to identifiers for all of those categories. Another categorization of interest to researchers is the particular methodological approach brought to the topic. It is observed that many papers bring more than one research methodology to bear, so each paper is keyed simultaneously to all the methods used therein . For example, Jintae Lee's (2003) paper on P2P

system user behavior employs qualitative methods to generate relevant factors, then uses those factors in a survey instrument to collect quantitative data. Lee's paper is thus entered in the database with relations to 'Empirical', 'Qualitative', 'Quantitative', 'Behavior', 'Ethics', and 'Outside the Firm'.

The result is a meta-framework, or a framework of frameworks, which, because it is implemented as a relational database, can be utilized to search on any axis or axes which are the current focus of interest. For example, an IS researcher with an interest in User Behavior can immediately query the database for all Behavioral studies of P2P filesharing systems and issues. He or she might then also sub-group those Behavioral studies by the system under consideration (Napster, Gnutella, Freenet, e.g.), by the methodology (Analytic, Empirical-Quantitative), or by any other keys implemented in the database. Conversely, one might want to see all of the Quantitative research on P2P systems; this set could then be further segmented into subsets such as 'Information Retrieval', 'Digital Rights Management', 'Behavior', and so on. Those subsets might even be further sub-grouped by whether they are 'Inside the Firm' or 'Outside the Firm', or again, by any of the other available keys.

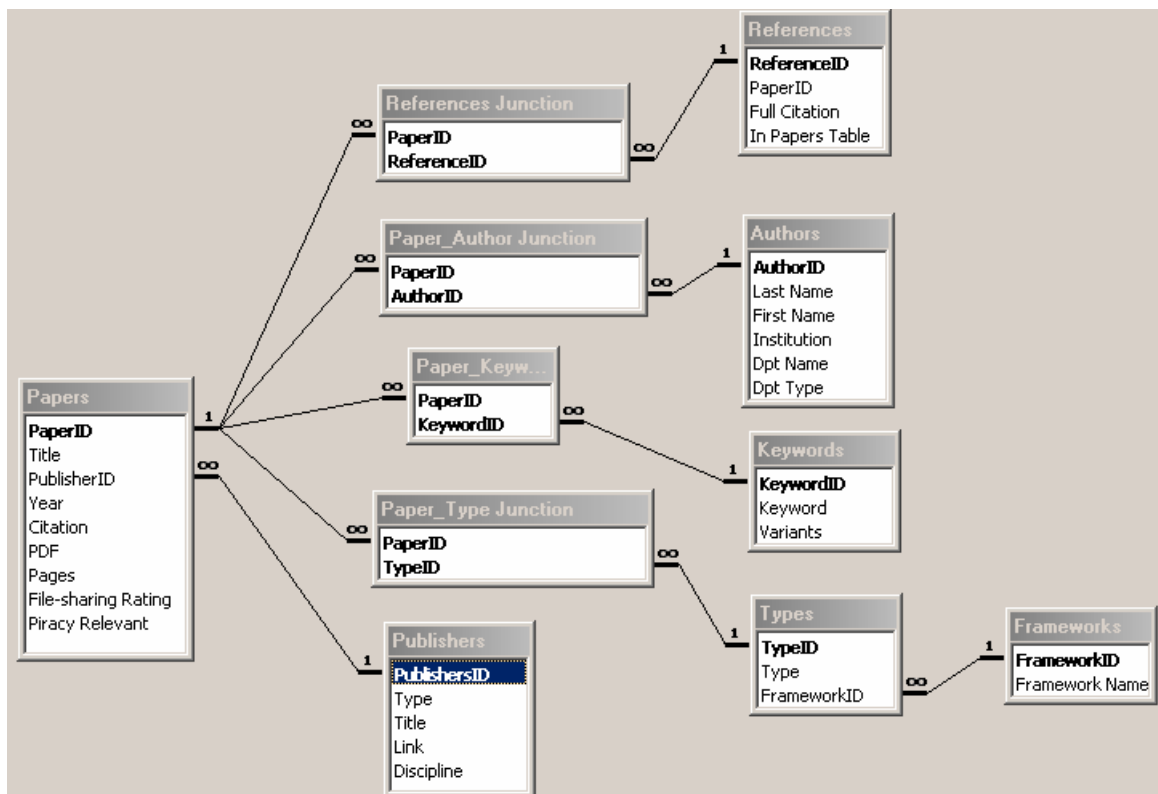
This approach allows for a highly flexible approach to searching the literature, since the database accomodates multiple approaches to categorization. It also highlights the inherently trans-disciplinary nature of the P2P filesharing literature, since so many papers are categorized simultaneously into disparate fields, both within and without IS, and makes explicit the links between the different fields which appear in the literature. Since these links are implemented as many-to-many relationships, the database design is open-ended; additional frameworks for categorizing the literature can be added as desired simply by adding new sets of keywords. The

section of the paper which follows describes the database which implements the meta-framework approach.

## 2.1 P2P Citations Database - Structure

'Citations.mdb' is a Microsoft Access database which categorizes 119 peer-reviewed articles on P2P filesharing systems from journals and conferences through April of 2004. Books (Oram, 2001; Fisher, 2004; Günnewig et al, 2003; Lessig, 2001; Litman, 2001; Gordon & Watt, 2003) working papers (Litman, 2003; Peitz and Waelbroeck, 2003), and some technical reports, such as the Pew Internet Reports which give numbers of users and files downloaded, (Graziano, 2001), while often of considerable interest, are not included. The basic structure can most easily be grasped through inspection of the Relationships layout in Figure 2:

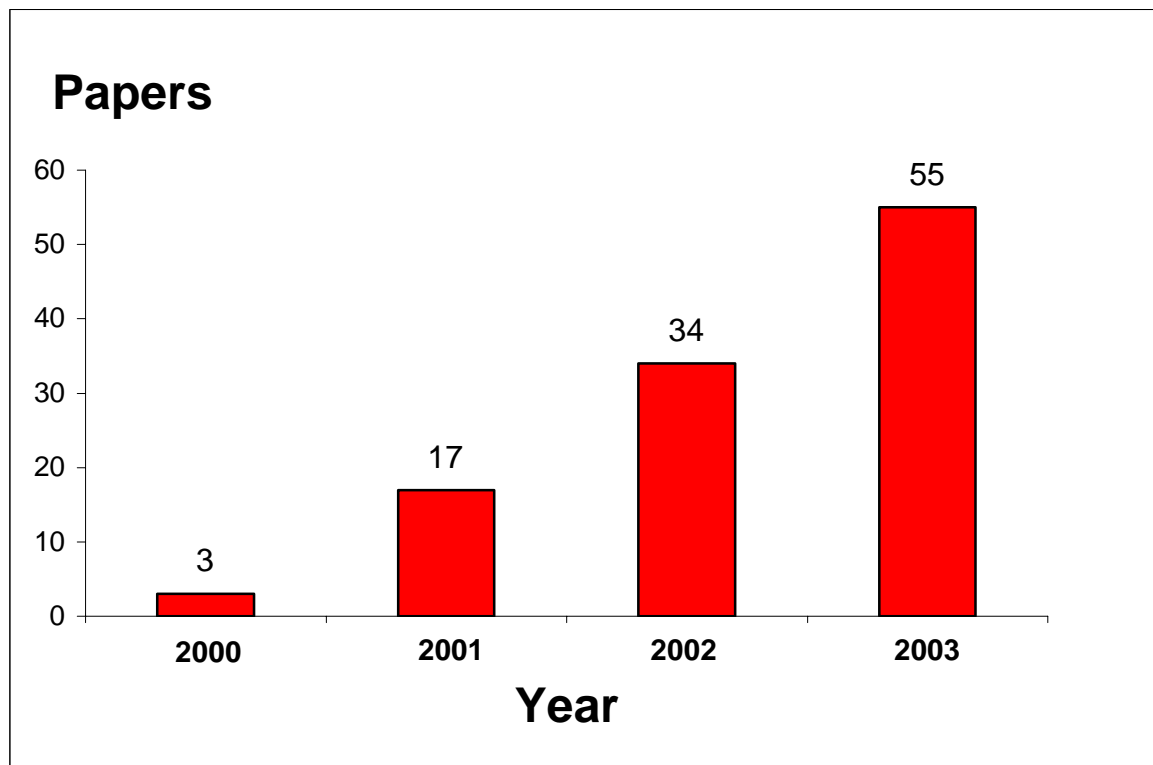
**Figure 2 - Relationships Structure of the Database**



On the left is the 'Papers' table, which includes all of the basic citation information, along with a field 'PDF' which provides a link to the actual paper, and two additional fields, 'File-sharing Rating' and 'Piracy Relevant', which describe the paper's relationship to issues of sharing media content, especially content which is under copyright. A further discussion of the uses of these fields will be given later in the paper. In the center of the Relationships window are all the junction tables, plus the 'Publishers' table, which includes a category for 'Type' (journal, conference) and for the 'Discipline' (IS, CS, Law, etc.). On the right are the 'References' table, which implements a cross-index, the 'Authors' table, which includes the institution and department from which the paper(s) originated, and the 'Types' and 'Keywords' tables, which are explained below. There is also a 'Frameworks' table on the far right, which groups the different Types.

Interest in P2P filesharing systems began with Napster. Cartwright (2000) noted that the groundbreaking P2P filesharing system lacked a viable business model, despite its considerable success, and suggested ways in which Napster's popularity might be translated into revenues. In the IS research community and elsewhere, interest has continued to build steadily, as a graph of the results of a query on paper counts by year shows (Figure 3):

**Figure 3 - P2P Filesharing Studies**



A further 11 papers from 2004 are included in the database. It is difficult to authoritatively produce an assessment of the state of P2P filesharing systems development and use at any particular point in time, because the technology, and industry's responses to and uses of it, are evolving so rapidly. A notable feature of the P2P filesharing literature is the inclusion in a great many papers of citations from the news media, which is a published resource much closer to the cutting edge of P2P technology than are academic publications.

## **2.2 P2P Citations Database - Classification Frameworks**

At present, there are four different categorizations of the P2P filesharing literature in the database:

- 1) **IT Artifact** - the views of technology described in Orlikowski & Iacono (2001)
- 2) **Methodology** - research methods
- 3) **Perspective** - the particular aspect of P2P systems addressed
- 4) **Source** - where the P2P system originates

In addition to the above, the papers can also be sorted or selected by author, publisher, discipline, date, and specific P2P topics such as anonymity, Napster, BitTorrent, free-riding (downloading files, but not sharing any with the network), piracy, fault tolerance, and so forth. The entries under each categorization can be seen in Figure 4, which was generated by a report run on the database:

**Figure 4 - Frameworks with Associated Types**

## *Frameworks1*

<i>Framework Name</i>	<i>TypeID</i>	<i>Type</i>
<i>IT Artifact</i>		
	101	Tool
	102	Proxy
	103	Ensemble
	104	Computational
	105	Nominal
<i>Methods</i>		
	201	Analytic
	202	Theoretical
	203	Metrics
	204	Prototype/Design
	205	Empirical
	206	Qualitative
	207	Quantitative
	208	Simulation
	209	Literature Survey
<i>Perspectives</i>		
	301	Economics
	302	Strategy
	303	Culture
	304	Law
	305	Digital Rights Management
	306	Information Retrieval
	307	Ethics
	308	Behavior
	309	Performance
	310	Human Computer Interaction
	311	Marketing
<i>Source</i>		
	401	Inside Firm
	402	Outside Firm

### **2.2.1 P2P Citations Database - Classification Frameworks - Inside or Outside the Firm?**

This most basic distinction arises from the fact that some work on P2P filesharing systems is being done to evaluate their suitability for the realization of institutional goals, while the rest addresses issues which arise in P2P systems which do not have their origin in institutions. For example, Schoder and Fischbach (2003) examine the current state of P2P technology and conclude that before it can be put to use inside the firm, characteristics such as network control, security, interoperability, information search and cost-sharing (limiting free-riding) need to be improved; there is general agreement on this point (Ratnasamy et al, 2001; Golle et al, 2001; Loo, 2003). There is also an appreciation of the power which P2P systems can bring to bear upon business-oriented tasks such as sharing data. Jensen and Xiao (2001) find that P2P systems can be an efficient means for distributing customized financial reports; Lu and Callan (2003) demonstrate that their content-based resource selection query model for a P2P distribution network can provide efficient retrieval of documents. Kant & Mohapatra, (2001) and Bawa et al (2003) demonstrate the feasibility of P2P for delivering web content. Izal et al (2004) show that P2P systems have potential to help relieve the burden on centralized servers which provide software updates for users. Papers were classified as "Inside the Firm" when they explicitly examined such business-oriented goals for P2P filesharing systems. An excerpt from a report which counts the number of papers in each category shows that such studies are in the minority (Figure 5):

**Figure 5 - Inside or Outside the Firm Counts**

<i>FrameworkID</i>	<i>First Of Type</i>	<i>CountOfPaperID</i>
<b>04</b>		
	<i>Inside Firm</i>	26
	<i>Outside Firm</i>	94

This disparity is likely due to the fact that P2P systems which are 'Outside the Firm' have been in place and functioning with millions of users since 1999, while business P2P systems are still in development. and thus not available for study in a corporate environment. While the technology of such P2P systems as Kazaa or WinMX is not mature from a corporation's or computer scientist's point of view--these systems are widely criticized in the Computer Science literature for the inefficient method of distributing search queries employed--it is at least mature enough to drive widespread adoption among Internet users. There are a lot of systems to study, and many performance issues to address. A query run on items in the Keywords table counts the number of papers studying specific systems and architectures (Figure 6):

**Figure 6 - Filesharing Applications for Digital Media**

## *Napster and others*

<i>Keyword</i>	<i>CountOfPaperID</i>
<i>BitTorrent</i>	1
<i>DirectConnect</i>	1
<i>eDonkey</i>	1
<i>FastTrack</i>	1
<i>Freenet</i>	4
<i>Gnutella</i>	21
<i>Kazaa</i>	4
<i>Morpheus</i>	1
<i>Napster</i>	18

There is a wide range of topics to be addressed on P2P systems Outside the Firm. Portmann & Seneviratne (2003) create a more efficient search method which exploits the power-law characteristics of P2P networks called Rumor-mongering. Landau (2002) reviews the legal situation and finds that, while current law makes filesharing of copyrighted material unambiguously illegal, the most prudent course for companies may be to pursue technological solutions to piracy. Klein et al (2002) show that P2P filesharing systems can reduce the economic value of copyrights. Heverly (2003) considers information goods as property in the digital age, and proposes that we create a new category to take account of the effects of technologies such as P2P systems which he calls the "Information Semicommons", which would strike a balance between traditional notions of property and completely unfettered distribution of

information goods. Samuelson (2004) holds that the non-infringing uses possible with P2P filesharing tools means that there is no legal right to generally prevent their use. Fetscherin, & Schmid (2003) and MacKenzie & Goldman (2000) both examine ways of defending against the threats posed by information goods piracy that occurs in P2P. Fetscherin, & Schmid find that different industries employ Digital Rights Management protections in different ways, and have differing degrees of belief that DRM can be effective. MacKenzie & Goldman take a different approach, emphasizing the importance of education in affecting users' behavior. Calkins (2002) identifies some of the ethical issues raised by the existence of open P2P filesharing systems such as Kazaa and Napster. Bhattacharjee, Gopal and Sanders (2003c) collect extensive data on the characteristics and beliefs of P2P users, and finds, among other things, that the sound quality of music files is not important to most users.

In contrast with the studies of P2P filesharing systems categorized as Inside the Firm, excepting only the fact that there is much room for improvement in technical performance, there is not a clear consensus on most of these points. This is the most salient difference between the two categories. It may arise from the fact that researchers as a group have a great deal of experience evaluating and contextualizing the factors relevant to computer systems which are designed and implemented for businesses, and very little in understanding systems which are emergent outcomes of forces outside the firm. The systems outside the firm may also simply be inherently more complex, since what Orlikowski and Iacono (2001) call the ensemble, or web of forces influencing the development and use of the IT artifact, is so much broader and intricately woven than the relatively controlled environment in which business applications are realized. To take one obvious example, any survey of users of P2P systems such as Kazaa is more likely than not to be an inquiry about actions which have been defined in some cases as illegal--and this is

only one of many complicating factors. This breadth and complexity is one of the main factors which motivated the development of a multi-dimensional means of categorization described in this paper, since it is not clear that any single preferable way of framing the issues in P2P filesharing literature exists.

### **2.2.2 P2P Citations Database - Classification Frameworks - Methods and Approaches**

Another way of segmenting the literature is by the research techniques employed. Most of the categories under this heading in Figure 4 are assumed to be self-explanatory. Papers were classified as Theoretical if they developed specific models of performance, process, or behavior. Papers which were explicitly described by their authors as "analysis", or attempted a logical analysis of relevant factors, whether formally, as mathematics, or informally, as discussion, were classified as Analytic. Figure 7 below reveals that the representation of P2P literature in these categories is quite lopsided:

Figure 7 - Methodologies and Research Approaches

## *Paper Counts by Type*

<i>FrameworkID</i>	<i>First Of Type</i>	<i>CountOfPaperID</i>
02		
	<i>Analytic</i>	64
	<i>Empirical</i>	34
	<i>Literature Survey</i>	1
	<i>Metrics</i>	13
	<i>Prototype/Design</i>	51
	<i>Qualitative</i>	4
	<i>Quantitative</i>	31
	<i>Simulation</i>	31
	<i>Theoretical</i>	14

This side-by-side comparison brings at least four points to our attention, which cast some light generally on the nature of studies in P2P filesharing systems. The first thing to note is that the greatest number of papers is observed in the Analytic category. This segment reflects the exploratory character of much research in P2P systems, as researchers wrestle with the problems of what questions to ask, how to frame the issues. Second, we also find high numbers of Prototype, Quantitative and Simulation studies, which reflect to a considerable extent the activity of computer scientists pushing forward the development of P2P systems and testing their innovations. Third, notice the gap under 'Literature'. The sole representative here (Foster, 2003),

is a literature review, from the point of view of this paper, only incidentally, since Foster's primary focus is on Grid Computing; it appears in this review not because of its coverage of P2P filesharing studies, but because it includes some comments about filesharing systems such as Gnutella for indications of what is already occurring in P2P distributed computing systems. Finally, Figure 7 indicates that Qualitative research is also in its infancy here, with just four papers in the list; two of these are by the same author (Giesler 2003a and 2003b), published in a Marketing journal.

### **2.2.3 P2P Citations Database - Classification Frameworks - Perspectives**

The categories under this framework were obtained by inspection of the extant literature, as the authors developed one or several of these fields as the focus of their studies of P2P filesharing systems. Often, classification could proceed directly from the title and/or keywords provided by the authors. Some types in this framework are areas within IS: Behavior (of users), Information Retrieval, Performance (of systems), and Human Computer Interaction. Some areas important to IS in the past have yet to turn up: IT and Organization, or IT investment and returns, for example, may appear later on, if businesses choose to develop P2P systems for their own use. Other categories represent reference disciplines to IS: Economics, Strategy (from Management), Culture (from Sociology), Law, Marketing, and Ethics (from Philosophy).

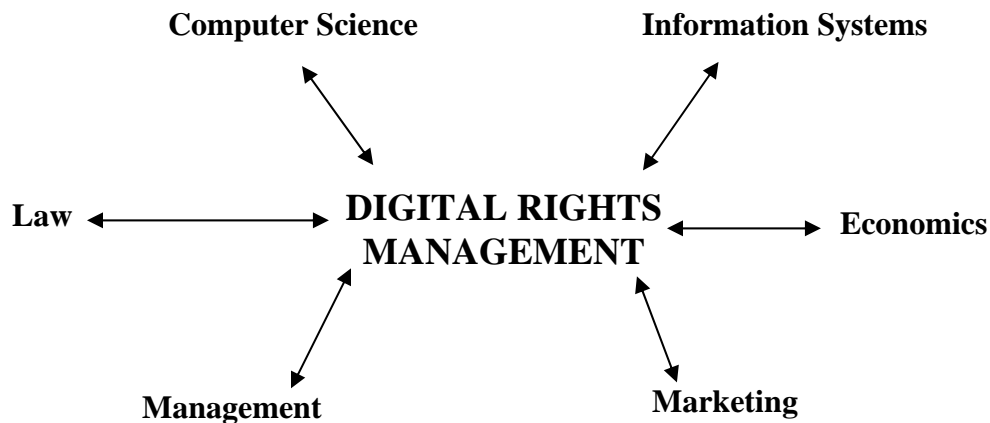
#### **2.2.3.1 P2P Citations Database - Classification Frameworks - Perspectives -**

##### **Digital Rights Management**

Digital Rights Management (DRM) is a major topic for P2P systems. It is also interesting as a classification problem, since DRM is a nexus where Computer Science, Law, Information

Systems, Management, Marketing, and Economics coincide (Figure 8). Because of its importance, and its explicit identification as the focus of interest for so many studies, it appears here as a separate category, with the understanding that DRM is inherently trans-disciplinary.

**Figure 8 - The DRM Nexus**



Some papers even make this point directly in the title, such as Samuelson's "DRM (and, or, vs.) the Law" (2003), or Garnett's "Digital Rights Management, Copyright, and Napster" (2001). Garnett here calls for policy makers and businesses to strike a balance between copyright protections and fair use; this argument is predicated on the assumption that sufficiently secure DRM technologies will become available, an assumption which is open to question.

In the P2P literature, there is in fact a lack of agreement in the P2P literature about what DRM can, and should do. Some analyses, like Garnett's, proceed upon the assumption that perfect or near-perfect control of digital media content via DRM will become available. Buhse (2001b) takes the differences of opinion on DRM into account in providing business strategy recommendations under varying conditions of piracy and DRM protection. Kwok (2002b) creates a license management model which attempts to satisfy both users and copyright holders

by providing both flexible use and copyright protection. Others, such as Gal et al (2003) or Fox & Wren (2001) hold that totally secure DRM is either difficult or impossible to achieve. These contradictory findings mirror a dichotomy of opinion within the media industries, where many music industry professionals do not believe that secure DRM can reduce piracy, while professionals in the film industry believe that it can (Fetscherin & Schmid, 2003). Similarly, some hold the opinion that strong DRM should be used to quash piracy (Tanaka, 2001) while others assert that to do so would impinge upon rights of fair use reserved to the public (Samuelson, 2003). Part of Tanaka's justification for his recommendation to resort to technological means to protect information goods is that legal enforcement on millions of users is not practical:

*"..., it is a daunting task for content providers to take appropriate and effective measures to ensure that the public does not illegally exploit their copyrights using P2P networks. Given the difficulties in legal enforcement, the most practical way to solve problems that were spawned by digital technology may be to utilize digital technology itself to solve the problem."*

The recording industry apparently arrived at precisely the opposite conclusion, after abortive attempts to develop secure copy-protected music media:

*"Major music labels, including Sony and Universal Music, have begun selling the "copy-proof" discs as a means of tackling the rampant spread of music piracy, which they claim is eating into sales. The new technology aims to prevent consumers from copying, or "burning," music onto recordable CDs or onto their computer hard drives, which can then be shared with other users over file-sharing Internet services such as Kazaa or Morpheus MusicCity. On Monday, Reuters obtained an ordinary copy of Celine Dion's newest release "A New Day Has Come," which comes embedded with Sony's "Key2Audio" technology. After an initial attempt to play the disc on a PC resulted in failure, the edge of the shiny side of the disc was blackened out with a felt tip marker. The second attempt with the marked-up CD played and*

*copied to the hard drive without a hitch."*—Wired News online, May 20, 2002, available at <http://www.wired.com/news/technology/0,1282,52665,00.html>.

Having discovered, at least provisionally, that effective DRM was not yet available, record companies apparently decided that in order to address piracy on the user's end of the distribution stream, it would be necessary to pursue legal means (copyright infringement lawsuits) of stopping P2P filesharing of copyrighted content among millions of users. Gal et al (2003) concluded that this confusing state of affairs would remain so for indefinite period:

*"Clearly the world isn't going to transition neatly to any one winning scenario; the reality is that the world is fragmented, and will remain that way for an extended period of time. The dynamics of consumer behavior, legal decisions, and technology improvements will guarantee that. Many millions of people will continue to purchase CD's, and there will almost certainly be a mixture of legal services going forward that will continue to compete with 'free: music available through other services (if the consumer is willing to take the risk to save money)."*

### **2.2.3.2 P2P Citations Database - Classification Frameworks - Perspectives -**

#### **Computer Science and Performance**

Computer Science is a reference discipline for IS, but does not appear as a category in the framework of 'Perspectives', because CS is already used both as a type classification in the database structure for journals/conferences, as well as for the departments to which the authors belonged at the time of publication. Searches for CS papers can thus always be obtained by queries on those fields. In addition, several of the categories already mentioned here are closely tied to CS: Information Retrieval, Performance (of systems), and Digital Rights Management. Papers with a CS focus (what Orlikowski and Iacono refer to as the 'Computational' view) are numerous in the P2P literature, so at this level of categorization, terms which are more specific

within CS are more useful than just 'CS' itself, which in any case is already implemented by other features of the database design.

The literature on Computer Science aspects of Peer-to-Peer systems is extensive, and not all of it pertains directly to an IS-oriented perspective. If we can envision Computer Science research running the gamut from pure science (such as quantum computing) to the development of user-ready applications, then the IS interest is closer to the application end of the spectrum. Consequently, papers from CS sources were included when they addressed application-level system implementation issues, or would be expected have a direct impact on such issues. Thus, the coverage of CS papers here is less deep than a CS-oriented review might prefer. In terms of breadth, the objective was to identify issues of interest to CS researchers concerning P2P system applications. Therefore, if a paper by a particular author addressing a particular issue was already present in the database, that was considered sufficient for the purposes of this survey. Adding all the additional papers by a particular author on a specific issue would not change any of the conclusions drawn herein about the literature as a whole, since the representation is already heavily skewed towards CS. Increasing the already high numbers of CS papers would simply make the gaps in the literature that much more salient.

A query of type counts again shows that the representation is somewhat lopsided (Figure 9):

**Figure 9 - Paper Counts by Perspective**

<i>FrameworkID</i>	<i>First Of Type</i>	<i>CountOfPaperID</i>
<b>03</b>		
	<i>Behavior</i>	27
	<i>Culture</i>	9
	<i>Digital Rights Management</i>	13
	<i>Economics</i>	18
	<i>Ethics</i>	4
	<i>Human Computer Interaction</i>	1
	<i>Information Retrieval</i>	16
	<i>Law</i>	11
	<i>Marketing</i>	7
	<i>Performance</i>	48
	<i>Strategy</i>	15

The largest category in Figure 9 is 'Performance', which reflects the activity of P2P system development taking place in the CS community. CS researchers value the enormous potential of P2P pooling of resources, and thus tend to concentrate on improving performance

beyond what has already been observed in existing architectures such as Napster and Gnutella.

To this end, CS professionals and researchers have several avenues to pursue. Kant and

Mohaptra (2001) give a detailed breakdown of points which need to be addressed:

*"Briefly, the most important issues include (a) devising efficient mechanisms for information location, (b) coping with network address translation and firewalls in providing interaction between peers, (c) intelligent searching and search response propagation mechanisms, (d) hybrid client-server and P2P approaches that can exploit the vast idle resources of P2P environment and yet provide the responsive and reliability of traditional client-server paradigm, (e) lightweight and nimble protocols to ensure good service to both the host and guest (i.e., P2P) applications on peers, (f) coping with various facets of hostile environment for transactional and real-time applications, and (g) performance characterization of P2P computing environment to enable comparative evaluation of many design choices."*

This list corresponds generally to the major topics found in the published literature, as seen

below in Table 2:

**Table 2 - Performance Issues**

<b><u>Major Issues</u></b>	<b><u>Papers</u></b>	<b><u>Findings</u></b>
<b>Fault Tolerance</b>	Stoica et al, 2003; Kapur et al, 2003; Fiat & Saia, 2002; Vishnumurthy et al, 2003; Koubarakis et al, 2003	These authors all agree that existing P2P systems are subject to loss of performance under certain conditions, and create prototypes to address the problems they identify.
<b>Free Riding</b>	Adar & Huberman, 2000; Saroiu et al, 2003	These authors find that free riding is prevalent on P2P networks, and that network performance is nevertheless fairly robust in the presence of significant free riding.
<b>Scalability</b>	Chawathe et al, 2003; Barbosa et al, 2004; Lv et al, 2002, Nejdil et al, 2003	These authors agree that the original Gnutella query process scales poorly, and suggest improved methods.
<b>Information Retrieval</b>	Yang et al, 2003; Klampanos & Jose, 2004	These authors agree that retrieval based on text tags (Gnutella) provides less useful searches than might be desired. Yang provides for content-based retrieval, while Klampanos et al creates a system which exploits network cluster structures.
<b>Anonymity</b>	Kubiatowicz, 2003; Clarke et al, 2001 and 2002; Serjantov, 2002; Freedman & Morris, 2002; Krishnan and Uhlmann, 2004	These authors agree that user anonymity is an important feature for some kinds of P2P systems. Kubiatowicz discusses design issues for anonymity, while the others propose, create, and test architectures for anonymity.

In addition to these primary topics, performance issues in general of all kinds are discussed. Lee et al (2002) and Trianafilou et al (2002) create prototypes to improve the speed of query resolution. Daswani & Garcia-Molina (2002), determine that load balancing could help protect against Denial-of-Service attacks. Sripanidkulchai et al (2002) and Krishnamurthy et al (2001) exploit emergent P2P network structures to improve performance. Liu & Kwok (2002) create a protocol to allow cross-communication between different P2P systems. Castro et al (2004) and Cohen et al (2003) create prototypes which can make more efficient usage of bandwidth and processing resources on P2P networks.

Fault Tolerance and Data Integrity are interesting problems, since these are characteristics without which 'legitimate' business-oriented P2P systems cannot be adopted. Fault tolerance is the ability to access data resources under conditions of, for example, node failure or communications interruptions (a full exploration of this topic is beyond the scope of this paper). In view of the possibility of malicious action on a P2P system, data reliability is treated by some authors as a contributing factor in fault tolerance. Business users would like to know that network performance will not be significantly diminished under conditions due either to technical fault or malicious behavior--yet the some of the same technologies which can guarantee fault tolerance are also those which could tend to nullify the attacks which some industries would like to make, and are making, against P2P filesharing networks used to exchange copyrighted content. There have already been instances of file 'spoofing' on the most broadly used networks--introducing a file with no content or useless content into the network

under the name of a widely pirated song in order to add friction to the file acquisition process, with the eventual goal of influencing the network towards complete collapse:

*"Madonna's label, Warner Bros. Records, has flooded the Web with fake MP3 files to thwart would-be tune thieves trying to access music off her new American Life before the album's released. Life isn't due to hit stores until next Tuesday. The decoy files have the appearance of full-length songs, and as such listeners don't know they've been had until they download and play the bogus tracks."*

--Lia Haberman for Eonline News, April 17, 2003.

<http://www.eonline.com/News/Items/0,1,11639,00.html>

. The same activity, spoofing, could present a real problem for the kinds of legitimate uses of P2P envisioned by CS application developers, so they are already developing features which will help guarantee the reliability of the data acquired over the P2P network. Gupta et al (2003) and Aberer & Despotovic (2001) both create and test systems for insuring that servants in a P2P network are trustworthy. Kamvar et al (2003) builds and tests an architecture which actually removes unreliable sources from the network.

Free Riding--making use of network content without contributing any of one's own in return--is prevalent on the P2P filesharing networks (Adar & Huberman, 2000), but has not brought about their collapse. Nor, according to the analyses conducted so far, is it likely to, since even the relatively primitive networks already in place can function well even when only a minority of members contribute (Ge et al, 2002). Some design work in progress now looks at ways of building incentive mechanisms for contributing into the P2P architecture itself, so that the threshold of free-riding beyond which network performance finally does degrade significantly is not approached (Ma et al, 2003, Cox & Noble, 2003).

### **2.2.3.3 P2P Citations Database - Classification Frameworks - Perspectives -**

#### **Human Computer Interaction, Ethics**

The lowest representation in Figure 9 occurs in Human Computer Interaction; the sole representative here (Good & Krekelberg, 2003) conducts a 'cognitive walkthrough' of the Kazaa interface, in order to develop an explanation of why some users appeared to be unaware that they were sharing files with the network. Here is a single design issue with trans-disciplinary implications: is the software switch to share or not share files from the user's computer distributed with a Default setting of 'ON' or 'OFF'? If users must opt-in to sharing, so to speak, then achieving a sufficient mass of sharing users to make the network self-sustaining may be difficult. But using a Default setting of 'ON' raises difficult issues of ethics, security, privacy, and legal liability. Good and Krekelberg found that the interface of the Kazaa application they used did not at all lend itself to a clear understanding of whether, how, or what users could share to the network from their own hard drives, and created a set of guidelines for design which address these issues.

Ethics also seems under-represented, given the highly controversial nature of P2P activity and the apparent huge cognitive disconnect between business (notably the RIAA) and many users' perspectives on the problem. In the context of a broader discussion which also looks at information filtering by portals and web-based consumer profiling, Calkins (2002) briefly discusses the ethical issues raised by Napster and identifies some of the unresolved questions business ethicists need to address. The other papers in this category only mention the ethics of P2P filesharing systems as a relevant factor in the course of a research program whose primary focus is elsewhere. Chen and Png (2003) include a distinction between ethical and unethical users in an economic model of information goods pricing and copyright. Good and Krekelberg (2003) conduct a cognitive walkthrough of the Kazaa interface. They find that many users have difficulty identifying precisely what is being shared on their platforms, with many users exposing

data they clearly intended to keep private, and conclude that designers of P2P applications have a responsibility to make the filesharing choices and defaults of the software more transparent to the user. Lee (2003) includes the category *Supports only legal files* in a survey-based study which characterizes users of P2P systems across a wide range of constructs. The publishing activity which directly addresses the most controversial issues raised by P2P filesharing systems is more concentrated in the Law category.

#### **2.2.3.4 P2P Citations Database - Classification Frameworks - Perspectives -**

##### **Law**

The numbers in Figure 9 for papers under 'Law' should not be considered meaningful here for the purposes of comparison to other categories. All papers in IS journals or conferences which addressed issues of law were included, but this does not exhaust the topic. The legal literature on copyright with significance for P2P filesharing systems is already vast, and not all relevant papers from law journals were sought out for inclusion. Instead, the list of legal papers here should be considered merely representative. Netanel (2003) argues that the social costs of strict control of digital media content would be unacceptably high, and suggests that a scheme of levies through compulsory licensing be instituted instead. Hosein et al (2003) holds that the legal attack on Napster had the unwanted effect of spawning new and more legally resistant technologies. Greene (2001) argues that the recording industry's legal attack on Napster represents a significant missed opportunity to take advantage of a powerful new distribution medium. Samuelson's (2003, 2004) views on DRM and the law have already been discussed above, as have Klein's (2002) on the law and P2P economics, and Heverly's (2003) on law and information goods.

The legal community, as represented in the P2P literature, is far from consensus on the topic of P2P filesharing systems. The issue of copyright alone is a muddy one, and adding P2P and DRM technologies to the mix only complicates things further. Copyright law is not only about protecting intellectual property. Its purpose, as enshrined in the U.S. Constitution itself, is to promote intellectual and creative activity, a principle which has in their past been realized through 'fair use' and the expiration of copyright. The legal balance between these interests as they apply to digital media distribution has swung towards property rights, a development due in part to the passage of the Digital Millenium Copyright Act (DMCA) (Clark, 2002; Mulligan, 2003), while the facts on the ground have conferred enormous power on users to avoid copyright restrictions (Lam & Tan, 2001). Some argue that existing copyright law, including the DMCA, is entirely inadequate to address the issues raised by P2P systems, and that new approaches which take the unique characteristics of information goods and technologies must be found (Stallman, 2002; Lessig, 2003).

#### **2.2.3.5 P2P Citations Database - Classification Frameworks - Perspectives -**

##### **Behavior**

The Behavior studies tended to fall into one of two main groups: 1) those behaviors which tend to effect system performance (Gummadi et al, 2002; Saroiu et al 2001 and 2002; Vaucher et al, 2002; Shneidman, 2003; Yang & Garcia-Molina, 2002), especially Free-Riding (Fig. 10), or 2) behaviors associated with piracy (Fig. 11). A query with two criteria captures the relevant papers:

**Figure 10 - Free Riding Behavior**

<i>Type</i>	<i>Keyword</i>	<i>Citation</i>
<b>Behavior</b>	<i>free riding</i>	<p>Horne, B., Pinkas, B. &amp; Sander, T. 2001. Escrow services and incentives in peer-to-peer networks. In Proceedings of the 3rd ACM Conference on Electronic Commerce. pp. 85-94.</p> <p>Cox, L. &amp; Noble, B. 2003. Samsara: Honor Among Thieves in Peer-to-Peer Storage. In Proceedings of the 19th ACM Symposium on Operating Systems Principles. pp. 120-132.</p> <p>Feldman, M., Lai, K., Chuang, J. &amp; Stoica, I. 2003. Quantifying Disincentives in Peer-to-Peer Networks. In Workshop on Economics of Peer-to-Peer Systems, Berkeley, California, June 2003.</p> <p>Ma, R., Lee, C., Lui, J. &amp; Yau, D. 2003. Incentive P2P networks: a protocol to encourage information sharing and contribution. ACM SIGMETRICS Performance Evaluation Review. September, vol. 31, #2, pp. 23-25.</p> <p>Ranganathan, K., Ripeanu, M., Sarin, A. &amp; Foster, I. 2003. To Share or Not to Share: An Analysis of Incentives to Contribute in Collaborative File Sharing Environments. Workshop on Economics of P2P Systems, Berkeley, California, June 2003.</p> <p>Sarciu, S., Gummadi, K. &amp; Gribble, S. 2003. Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts. Multimedia Systems. August, vol. 9, #2, pp. 170-184.</p> <p>Krishnan, R., Smith, M., Tang, Z. &amp; Telang, R. 2004. The Impact of Free-Riding on Peer-to-Peer Networks. In Proceedings of the 37th Hawaii International Conference on System Sciences</p> <p>Becker, J. &amp; and Clement, M. 2004. The Economic Rationale of Offering Media Files in Peer-to-Peer Networks. In Proceedings of the Hawaii International Conference on System Sciences.</p> <p>Adar, E. &amp; Huberman, B. 2000. Free Riding on Gnutella. First Monday. October, vol. 5, #10.</p>

**Figure 11 - Piracy Behavior**

<i>Type</i>	<i>Keyword</i>	<i>Citation</i>
<b>Behavior</b>	<i>piracy</i>	<p>Horne, B., Pinkas, B. &amp; Sander, T. 2001. Escrow services and incentives in peer-to-peer networks. In Proceedings of the 3rd ACM Conference on Electronic Commerce. pp. 85-94.</p> <p>Bhattacharjee, S., Gopal, R. &amp; Sanders, G. 2003. Digital music and online sharing: software piracy 2.0? Communications of the ACM. July, vol. 46, #7, pp. 107-111.</p> <p>Graham, L. 2003. RIAA lawsuits appear to be a "win" for the record industry, but winning back the hearts of consumers is another matter. Nov. 5th, <a href="http://www.npd.com/press/releases/press_031105.htm">http://www.npd.com/press/releases/press_031105.htm</a></p> <p>Bhattacharjee, S., Gopal, R., Lertwachara, K. &amp; Marsden, J. 2003. No More Shadow Boxing with Online Music Piracy. Strategic Business Models to Enhance Revenues. In Proceedings of the 36th Hawaii</p> <p>Becker, J. &amp; and Clement, M. 2004. The Economic Rationale of Offering Media Files in Peer-to-Peer Networks. In Proceedings of the Hawaii International Conference on System Sciences.</p> <p>Strahilevitz, L. 2003. Charismatic Code, Social Norms, and the Emergence of Cooperation on the File-Swapping Networks. Virginia Law Review, vol. 89.</p>

Notice that these two perspectives are at cross-purposes, as far as copyright issues are concerned. The papers in Fig. 10 mostly seek by various means to reduce free riding, in order to increase the efficiency, reliability and robustness of the network; such efficient networks would then become even more powerful and usable tools for sharing digital media content, including pirated content. The papers in Fig. 11, on the other hand, are looking at dealing with the possibly negative consequences of good network performance. The papers on piracy do generally agree that this is a problem that will likely not be solved anytime soon, and suggest various strategies for dealing

with it, including the adoption of new business models such as the use of compulsory licenses (Landau, 2002; Netanel, 2003).

Other topics of interest under 'Behavior' include how users connect and disconnect from the network (Douceur, 2003); identification of the factors, such as price, speed, reliability and stability, which influence user adoption of a P2P system (Lee, 2003; Strahilevitz, 2003); how servers (machines on the P2P network from which files are downloaded) are chosen (Aberer & Despotovic, 2001), and consumption behaviors online (Giesler & Pohlmann, 2003a and 2003b; Yang et al, 2003).

#### **2.2.3.6 P2P Citations Database - Classification Frameworks - Perspectives -**

##### **Economics**

The studies from the 'Economics' perspective can all be classified as addressing one of two concerns:

- 1) Analysis of conditions which affect network performance (Golle et al, 2001; Becker & Clement, 2004; Turner & Ross, 2004; Ranganathan et al, 2003; Lai et al, 2003; Feldman et al, 2003; Krishnan et al, 2004)
- 2) Analysis of business conditions under the threat of piracy (Chen & Png, 2003; Bhattacharjee et al, 2003b; Snir, 2003; Horne et al, 2001; Klein et al, 2002)

Notice that Economics papers divide in the same way as does much of the Behavior literature. They are about either P2P network performance or network piracy, and thus are also at cross purposes from the point of view of copyright issues. The lessons learned from the

performance analyses are intended to make P2P systems more powerful and efficient, and would therefore tend to increase their effectiveness in exchanging content, including pirated content.

The analyses of business conditions are not in complete agreement. Snir's analysis, and to a lesser extent, Klein's, indicate that the RIAA should attempt to shut down online piracy, while both Chen and Bhattacharjee find that such a strategy can be sub-optimal for industry revenues under some conditions. The differences of course have their origin in the different assumptions made in the models. Snir's model, for example, assumes that the perceived quality of online music obtained from P2P filesharing systems is inferior to that of purchased music, but there is empirical data which indicates that this assumption may not be warranted (Bhattacharjee et al, 2003c). One difficulty in the economic analysis of information goods distributed on P2P filesharing systems is the lack of traditional supply and demand relationships. The nature of the P2P resource is to cause a file which is in high demand to become more, not less available. In order to be useful, economic analyses need to take the peculiar characteristics of information goods (Clemons & Lang, 2003; Chen & Png) into account when models are built.

#### **2.2.3.7 P2P Citations Database - Classification Frameworks - Perspectives -**

##### **Strategy**

The 'Strategy' category papers are nearly all alike in one way, and dichotomous in another. A discussion of P2P systems from the point of view of a business hoping to deploy P2P filesharing technology to achieve 'legitimate' business goals appears in Kwok, Lang & Tam, (2002a), which presents a 3-layer P2P services model defining performance measures and success factors for P2P implementations Inside the Firm. Many papers here are about responses to the challenges posed by P2P filesharing systems from Outside the Firm. The recommended

strategic responses to P2P systems like Kazaa, then, generally fall into one of two categories: 1) fight back (Premkumar, 2003; Easley et al, 2003); or 2) concede, at least partially, control of distribution to P2P systems (Buhse, 2001a; Fox, 2001; Davis, 2003). A few papers address both strategies (Clemons, Gu & Lang, 2002; Buhse, 2001b). While the RIAA through 2004 has chosen the former approach, it may be that a revenue-maximizing strategy does not necessarily involve the elimination of online piracy (Bhattacharjee et al, 2003a).

### **2.2.3.8 P2P Citations Database - Classification Frameworks - Perspectives -**

#### **Culture, Marketing**

The relatively low numbers of papers under 'Culture' or 'Marketing' in Figure 9 suggest that there is room for contributions in these areas as well. Giesler (2003a and 2003b) uses qualitative analyses to study the culture of filesharing communities, and finds that P2P users tend to define themselves in opposition to the music industry, and characterizes network populations as gifting communities. The latter approach solves one of the confounding properties of P2P networks, which is that exchanges of goods take place without an explicit, reciprocal economic transaction involved. In the extremely unlikely event that two members of a P2P network simultaneously traded files, such an occurrence would be merely coincidental in its timing, and would not indicate the existence of any economic characteristics normally associated with barter. The prevalence of free riding on the P2P networks highlights the point that the behaviors observed there are not directly linked to notions of exchanging value for value. This is in contrast to users' perceptions of resources whose use by others reduces the amount available by the user: some P2P participants will deliberately under-report their bandwidth to the network in order to minimize its use by others:

*"Note that as high as 30% of the users that report their bandwidth as 64 Kbps or less actually have a significantly greater bandwidth. In Napster (and any similar system), a peer has an incentive to report a smaller bandwidth than the real value, in order to discourage others from initiating downloads and consuming the peer's available bandwidth"*  
(Sarioiu et al, 2003)

This dichotomy in user behavior--generous to the point of illegality on the one hand, selfish and deceptive on the other--highlights the central economic problem created by P2P networks. When users 'give' their music away on the network, they do not personally lose any value in their own music collection at all, as they would if they were giving away music CD's. The economic incentives not to share have to do only with the CPU cycles and bandwidth consumed, and the relatively basic P2P architectures employed have proven to be relatively robust in the face of self-interested hoarding of those resources, even by a majority of participants. The incentives to participate, on the other hand, are strong: access to music which they would otherwise have to purchase, or which might no longer be available through traditional channels at any price (Bhattacharjee et al, 2003). Copyright owners can attempt to rebalance the equation by increasing the perceived cost of detection of illegal user behavior (Chen & Png, 2003), but raising the cost sufficiently may be difficult, as we saw in the legal literature (Tanaka, 2001).

Strahilevitz (2003) interprets the Napster network as an arena in which social cooperative norms were successfully fostered by the system's creators, and suggests that these norms are the primary factor guiding user behavior:

*"If norms, and not the law, are what motivates consumers to act, then a wiser strategy for the RIAA and their allies might be to think about ways in which they might weaken the cooperative norms that have arisen among users of these networks."*

In this respect he is in agreement with Giesler's analyses; if P2P networks are in fact gifting communities, then it is reasonable to suppose that they could have particular values specific to the community. This notion is supported indirectly by some of the empirical data on P2P users, which suggests that P2P filesharing communities are defined in part by their demographic characteristics (Bhattacharjee, 2003c). Lechner & Hummel (2002) and Robinson & Halle (2002) take similar lines, analyzing P2P systems as communities. Unlike some of the other categories under the 'Perspectives' framework, there is at least general agreement in the 'Culture' research on P2P filesharing systems about how to proceed, and what the forces at work are. On the other hand, both the findings and recommendations are far from specific, and much work remains to be done before 'Culture' research can link up with, say, 'Strategy' or 'Economics'.

Of the papers we have looked at, those under 'Culture' are the least trans-disciplinary; these studies do not engage the IT artifact (specific P2P systems' architectures and interfaces) at all; it is merely mentioned as the basis upon which the cultural artifacts are realized--Orlikowski and Iacono (2001) refer to this approach as the 'Nominal' view of IT. Lechner and Hummel (2002) is a notable exception, in that it attempts to make a connection between system architectures and the social communities which result, concluding that virtual communities supported by P2P technologies such as Napster and Gnutella have the potential to create economic value.

#### **2.2.4 P2P Citations Database - Classification Frameworks - The IT Artifact**

Since the topic of this literature review is a particular class of IT artifacts, P2P filesharing systems, it was considered appropriate also to classify the papers according to the categories

described by Orlikowski and Iacono (2001). The following list explains how decisions were made to put a paper into one category or another.

**Tool View** - examines P2P systems with particular regard to the uses to which they are put, or the capabilities they provide. For example, Bouvin (2002) discusses the potential and challenges of using P2P systems for hypermedia, identifying trust and resource discovery in particular as problems which need to be solved.

**Proxy View** - looks at users' attitudes and beliefs about the P2P systems they use. For example, Lee's 2003 paper develops and uses a survey instrument on this topic. The quantitative data relevant to the proxy view as 'diffusion' (such as numbers of persons using a technology) tends to appear in sources which were not included in the database, such as the Pew Internet Research reports, which can collect and publish this swiftly changing data much more quickly than academic institutions.

**Ensemble View** - examines P2P systems as they are influenced by the larger contexts in which they function. For example, Lakshminarayanan & Padmanabhan (2003) analyze the impact of the bandwidth of users' connections to the P2P network on its performance, and find that the 'last-hop' upstream bandwidth was the most constrained resource. The distributed nature of P2P filesharing systems makes this category a topic of major interest, especially among CS researchers.

**Computational View** - looks at the algorithms actually used in the P2P systems. For example, Ripeanu et al (2002) analyze the emergent structure of Gnutella networks and use that information to create changes to the Gnutella architecture and protocols which will increase performance and scalability.

**Nominal View** - P2P systems are mentioned only as the basis for other topics which are the primary focus of the paper, without further discussion of any characteristics of the IT itself. Most of the Law and Culture literature takes this approach.

As we have seen in the other frameworks, the P2P representation under this method of classification is quite lopsided (Fig. 12):

**Figure 12 - IT Artifact Framework Distribution**

<i>FrameworkID</i>	<i>First Of Type</i>	<i>CountOfPaperID</i>
01		
	<i>Computational</i>	54
	<i>Ensemble</i>	24
	<i>Nominal</i>	21
	<i>Proxy</i>	4
	<i>Tool</i>	15

The high number for 'Computational' reflects the immature state of P2P systems, as CS researchers work to create improvements on existing architectures, and new architectures for new purposes. The low number for 'Proxy' suggests that there is potential for original contribution here. We can imagine that it would be useful to know, both quantitatively and

qualitatively, how new P2P systems propagate through the Internet and gain critical mass, or how network performance affects user perceptions of usefulness.

### **2.3 P2P Citations Database - Locating the Research**

In addition to the usual citation information, the database structure includes fields for naming and classifying the institutions and departments in which the authors worked at the time of publication, and for naming and classifying the publishing bodies. The following are available as values under 'Dpt Type' in the 'Authors' table (Fig. 13):

#### **Figure 13 - University Departments of the P2P Authors**

---

## *Authors*

---

### *Dpt Type*

---

*Business*

*Communications*

*Company*

*CS*

*Economics*

*Engineering*

*Finance*

*IS*

*Law*

*Management*

*Marketing*

*Sociology*

There is one value for this field in the database, 'Company', to capture the data on researchers who work in private businesses rather than universities. Again, the wide variety of departments contributing to the P2P filesharing literature illustrates the topic's trans-disciplinary nature. The following are available as values in the 'Discipline' field of the 'Publishers' table (Fig. 14):

**Figure 14 - Research Disciplines of the Publishers**

---

## *Publishers Disciplines*

---

---

### *Discipline*

---

*CS*

*Economics*

*IS*

*Law*

*Management*

*Marketing*

*Sociology*

Here the breadth of representation is not quite as great as in the Authors table, but the major reference disciplines for IS relevant to P2P are visible.

Some interesting searches can be run on the database with these supplementary fields. For example, a query on Papers with Institutions reveals the universities and labs where significant interest in P2P research exists. These include AT&T labs with 9 papers, Stanford University with 8, MIT with 4, and the University of California at Berkeley far in the lead with 15. Close inspection of the titles of the conferences shows that the P2P research community is becoming more organized, as the earlier papers are distributed among conferences of various types, while the later ones also begin to appear in workshops dedicated specifically to P2P systems. The most prominent journals to date are *Communications of the ACM*, which has published extensively on P2P filesharing systems, with 13 papers; IEEE has also taken a significant interest in P2P systems. Some sorting by hand would be required, but one might also use the database to distinguish between European and U.S. research, using the information in the field of the 'Authors' table labeled 'Institutions'.

## 2.4 P2P Citations Database - Filesharing Ratings

One cannot read this literature without noticing that some papers seem quite sympathetic to filesharing activity on P2P systems 'Outside the Firm' such as Kazaa, while others seek to find ways to reduce or eliminate this activity. Rather than pass over this implicit (and sometimes explicit) content in silence, it was decided to implement two extra fields for the Papers table, called 'Filesharing Rating' and 'Piracy Relevant'.

The 'Filesharing Rating' is a scale from 1 to 5, according to the following divisions:

- 1** - Explicitly calls for an end to filesharing of copyrighted content on P2P. For example, Tanaka (2001) reviewed the legal issues raised by Napster and said in conclusion:  
"...efforts to technologically enforce and eliminate unauthorized use will and should continue consistent with the various accommodating measures mentioned above. If not, content providers will lose the race against vicious pirates utilizing P2P and other newly emerging technologies." The unmistakable view of filesharing conveyed by this language results in this paper receiving a rating of '1'.
- 2** - Focuses on negative impacts of P2P filesharing or makes recommendations which if followed would be expected to reduce filesharing. For example, Snir's (2003) economic analysis indicates that "...closing music sharing services, suing ISPs and universities who are complacent about sharing, and "polluting" the pool of available music are all in the labels' best interest." However, Snir does not explicitly condemn filesharing per se, and thus receives a rating of '2'.
- 3** - Neutral or does not apply. There are studies which do not attempt to provide support, implicitly or explicitly for one side or the other of the filesharing controversy. For

example, the several studies by Saroiu and his colleagues (2001, 2002, 2003) focus on the collection of quantitative data on the behavior of P2P networks, noting that the purpose of the application is filesharing but neither condemning nor supporting this activity, and thus earning a filesharing rating of '3'.

**4** - Focuses on positive effects of P2P filesharing, or makes recommendations which if followed would be expected to improve filesharing. Most of the work under the 'Computational' category in the IT Artifact framework gets a '4', because the prototypes and algorithms explored are deliberately intended to improve the performance of P2P networks by providing, for example, more efficient search, better data reliability or load balancing, or by adding more features, such as anonymity. Since these improvements would increase the attractiveness of P2P applications for adoption by a wide base of users, these performance-oriented studies warrant a rating of '4'. Also included here are papers whose analyses point to some useful effects, from Outside the Firm, of P2P filesharing applications. For example, Bhattacharjee et al's (2003b) economic analysis draws the perhaps counter-intuitive conclusion that "under certain conditions, the seller's revenue maximizing position occurs in an environment with piracy activity above minimum levels."

**5** - Explicitly supports P2P filesharing, criticizes attempts to control it. For example, Fox and Wrenn (2001) recommend that the music industry transition to a free content 'broadcasting model' in which P2P systems are a primary mode of distribution, and thus receive a rating of '5'.

'Piracy Relevant' is a binary field in the Papers table, included so that papers which have no relevance to the piracy problem can be excluded from some queries.

With these ratings, one can segment the literature in different ways and look for contrasts. A side-by-side comparison of the averages for CS and IS papers seems to show a real difference:

**Table 3 - Filesharing Ratings by Discipline**

	<i>IS Journals Only</i>	<i>Computer Science</i>
<b>Behavior</b>	3	3.6
<b>Economics</b>	2	4
<b>Information Retrieval</b>	4	4
<b>Performance</b>	3.6	3.9

It is likely that part of this difference stems from the fact that CS researchers, by the very nature of the profession, are dedicated to improving the performance of computer systems. In this case, that means improving the efficiency and features of P2P filesharing systems. It is CS departments which are creating architectures to support anonymity, for example, though anonymous P2P systems could make the RIAA's legal strategy more difficult to carry out. Likewise, it is the nature of their discipline that IS researchers are concerned with solving problems for business. In the past this often meant solving problems in getting systems to work as desired, but in the case of P2P filesharing applications Outside the Firm, it may also mean helping them respond effectively to the challenges posed by such systems.

The filesharing ratings obtained here should not be considered a rigorous finding. To derive reliable numbers for this sort of analysis would of course require extensive further work.

An effort was made to make the definitions of the categories 1 through 5 above as simple and straightforward as possible, but confirmation of any findings would depend upon independent review by several different readers, and subsequent production of a number for inter-rater reliability which could give us confidence in the reality of the ratings.

### **3. Conclusion**

This paper has conducted a comprehensive review on P2P filesharing systems, and described a multi-dimensional framework, implemented in a relational database, which allows for flexible searches and classifications of the literature, depending upon the concerns of the researcher. Findings of the literature were summarized, according to four different frameworks.

The author has already found the database to be enormously useful, both in the preparation of this review, and in ongoing research projects (Hughes and Lang, 2004). The query capabilities make the recovery and organization of literature references for writing far more efficient than, for example, handling stacks of paper copies and printouts of articles, and give indications of where future research is indicated. The present intention is to continue to update the database, as further articles appear in journals and conferences.

Some immediate directions for P2P research are suggested by the broad overview which trans-disciplinary scope of the database provides. For example, the RIAA has pursued a range of strategies to protect copyright in dealing with the challenges posed by P2P filesharing systems; however, an examination of the literature, in particular as features are added to P2P tools, reveals that successful implementation of these strategies is by no means assured, as Table 4 illustrates:

#### **Table 4 - Industry Strategies Versus P2P Filesharing**

<b>RIAA Strategy</b>	<b>Corresponding Threat</b>
Digital Rights Management	Cracking security features (hackers) Re-recording to strip off DRM protection
Higher Quality Recordings (SuperAudioCD, DVD-A)	Sound quality not currently a decisive issue among MP3 users (Bhattacharjee et al, 2003c) Improved compression algorithms improve MP3 quality
P2P System Lawsuits (Napster shutdown)	New architectures without central indexing (Gnutella)
Lawsuits against Users	Anonymous P2P tools Distributed File locations (no specific uploader to identify)
Posting fake files to clog P2P networks	Development of P2P application features to identify server quality, remove unreliable data providers
User education initiatives to discourage filesharing	"Free information" culture fostered in P2P communities, implicit in P2P applications

It has been previously noted that relatively few studies have attempted to explore the impacts of cultural values on P2P systems, and vice versa. This is one gap in the literature which, if filled, may be able to provide new and useful insights into the evolution of P2P tools, their user communities, and the implications for the strategic responses of industries vulnerable to the challenges posed by P2P filesharing tools from Outside the Firm.

There appears to be immediate potential for additional work in the area of information economics as applied to P2P tools. The studies reviewed here make a variety of assumptions, not all of which are necessarily consistent with other studies, or with some of the quantitative data now becoming available. One salient example is the function of sound quality as a variable in models of users' decision-making under different conditions, as in Snir (2003). It was mentioned

above that quantitative data exists (Bhattacharjee et al, 2003c) which calls this assumption into question. It may be productive at this point to review some of the analytic modeling being done in light of the peculiar characteristics of information goods on P2P networks, and the facts we are beginning to learn about P2P user behaviors and values.

Overall, the picture is of a disruptive technology. Significant disagreement exists in many areas concerning the proper uses, impacts, and responses to this new kind of system, which arises outside of any institutional context, and operates outside of institutional control. The technology, though primitive in some ways, is extraordinarily powerful, and has drawn many enthusiastic friends, as well as determined foes. Further development of P2P applications is swift and ongoing, and promises to make P2P networks even more powerful and therefore more disruptive. Impacts have been felt most strongly in the music industry, but similar patterns are developing and are expected in movie and video media industries. Applications to share moving-image files, which are two orders of magnitude larger than music files, are already working and being widely used (BitTorrent). The RIAA has yet to develop any strategies in response to filesharing piracy which have not or cannot be thwarted by P2P applications already built or in development. Whatever the future of information goods holds, it is certain that P2P filesharing systems will play a central role.

## References

- Aberer, K. & Despotovic, Z. 2001. Managing Trust in a Peer-to-Peer Information System. In *11th International Conference on Information and Knowledge Management*. pp. 310-317.
- Adar, E. & Huberman, B. 2000. Free Riding on Gnutella. *First Monday*. October, vol. 5, #10.
- Alexander, P. 2002. Peer-to-Peer File Sharing: The Case of the Music Recording Industry. *Review of Industrial Organization*, March, vol. 20, #2, pp. 151-161.
- Barbosa, M., Costa, M., Almeida, J. & Almeida, V. 2004. Using Locality of Reference to Improve Performance of Peer-to-Peer Applications. In *Proceedings of the 4th International Workshop on Software and Performance*. pp. 216-227.
- Barki, H., Rivard, S. & Talbot, J. 1993. A Keyword Classification Scheme for IS Research Literature: An Update. *MIS Quarterly*, vol. 17, #2, pp. 209-226.
- Bhattacharjee, S., Gopal, R. & Sanders, G. 2003a. Digital music and online sharing: software piracy 2.0? *Communications of the ACM*. July, vol. 46, #7, pp. 107-111.
- Bhattacharjee, S., Gopal, R., Lertwachara, K. & Marsden, J. 2003b. Economic of Online Music. In *Proceedings of the 5th International Conference on Electronic Commerce*. pp. 300-309.
- Bhattacharjee, S., Gopal, R. & Sanders, G. 2003c. Digital music and online sharing: software piracy 2.0? *Communications of the ACM*. July, vol. 46, #7, pp. 107-111.
- Bouvin, N. 2002. Open Hypermedia in a Peer-to-Peer Context. In *Proceedings of the 13th ACM Conference on Hypertext and Hypermedia*. pp. 138-139.
- Buhse, W. 2001a. Digital Rights Management for Music Filesharing Communities. In *Proceedings of the Seventh Americas Conference on Information Systems (AMCIS 2001)*. pp. 1537-1543.
- Buhse, W. 2001. Categorizing Distribution Model Scenarios for Online Music Source. *Proceedings of the 2nd International Conference on Electronic Commerce and Web Technologies*. pp. 337-346.
- Calkins, M. 2002. Rippers, Portal Users, and Profilers: Three Web-Based Issues for Business Ethicists. *Business and Society Review*. Spring, vol. 107, #1, pp. 61-75.
- Cartwright, S. 2000. Napster: A Business in Search of a Model. *Journal of Business Strategy*, Sept/Oct, vol. 21, #5, pp. 28-32.
- Casanova, H. 2002. Distributed Computing Research Issues in Grid Computing. *ACM SIGACT News*. September, vol. 33, #3.

Castro, M., Costa, M. & Rowstron, A. 2004. Should We Build Gnutella On a Structured Overlay? *ACM SIGCOMM Computer Communication Review*. January, vol. 34, #31, PP. 131-136.

Chawathe, Y., Ratnasamy, S., Breslau, L., Lanham, N. & Shenker, S. 2003. Making Gnutella-Like P2P Systems Scalable. In *Proceedings of 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. pp. 407-418.

Chen, Y., & Png, I. 2003. Information Goods Pricing and Copyright Enforcement: Welfare Analysis. *Information Systems Research*, vol. 14, #1, pp. 107-123.

Clark, D. 2002. How Copyright Became Controversial. In *Proceedings of the 12th Annual Conference on Computers, Freedom and Privacy*. pp. 1-10.

Clarke, I., Sandberg, O., Wiley, B. & Hong, T. 2001. Freenet: A Distributed Anonymous Information Storage and Retrieval System in Designing Privacy Enhancing Technologies. *International Workshop on Design Issues in Anonymity and Unobservability*.

Clarke, I., Miller, S., Hong, T., Sandberg, O. & Wiley, B. 2002. Protecting Free Expression Online with Freenet. *IEEE Internet Computing*, January/February, vol. 6, #1, pp. 40-49.

Clemons, E., Gu, B. & Lang, K. 2002. Newly Vulnerable Markets in an Age of Pure Information Products: An Analysis of Online Music and Online News. *Journal of Management Information Systems*. vol. 19, #3, pp. 17-41.

Clemons and Lang, The Decoupling of Value Creation from Revenue: A Strategic Analysis of the Markets for Pure Information Goods. *Information Technology and Management*, vol.4, pp. 259-287.

Cohen, E., Fiat, A. & Kaplan, H. 2003. A Case for Associative Peer-to-Peer Overlays. *Computer Communication Review*. January, vol. 33, #1, pp. 95-100.

Conti, M. & Giordano, S. 2001. Mobile Ad-hoc Networking. In *Proceedings of the 34th Hawaii International Conference on System Sciences*.

Cox., L. & Noble, B. 2003. Samsara: Honor Among Thieves in Peer-to-Peer Storage. In *Proceedings of the 19th ACM Symposium on Operating Systems Principles*. pp. 120-132.

Damiani, E., Vimercati, D., Paraboschi, S., Samarati, P. & Violante, F. 2002. A Reputation-Based Approach for Choosing Reliable Resources in P2P networks. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*. pp. 207-216.

- Daswani, N. & Garcia-Molina, H. 2002. Query-flood DoS Attacks in Gnutella. In *Proceedings of the 9th ACM conference on Computer and communications security*.
- Davis, R. 2001. The Digital Dilemma. *Communications of the ACM*. February, vol. 44, #2, pp. 77-83.
- Douceur, J. 2003. Is Remote Host Availability Governed by a Universal Law? *ACM SIGMETRICS Performance Evaluation Review*. vol. 31, #3, pp. 25-29.
- Easley, R., Michel, J. & Devaraj, S. 2003. The MP3 Open Standard and the Music Industry's Response to Internet Piracy. *Communications of the ACM*. November, vol. 46, #11, pp. 90-96.
- Edwards, W., Newman, M., Sedivy, J., Smith, T., Balfanz, D., Smetters, D., Wong, H. & Izadi, S. 2002. Using Speakeasy for Ad Hoc Peer-to-Peer Collaboration. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*. pp. 256-265.
- Feldman, M., Lai, K., Chuang, J. & Stoica, I. 2003. Quantifying Disincentives in Peer-to-Peer Networks. In *Workshop on Economics of Peer-to-Peer Systems*, Berkeley, California, June 2003.
- Fetscherin, M. & Schmid, M. 2003. Comparing the Usage of Digital Rights Management Systems in the Music, Film and Print Industry. In *Proceedings of the 5th International Conference on Electronic Commerce*. pp.316-325.
- Fiat, A. & Saia, J. 2002. Censorship-Resistant Peer-to-Peer Content Addressable Networks. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms*. pp. 94-103.
- Fisher, W. 2004. *Promises to Keep: Technology, Law, and the Future of Entertainment*. Stanford University Press.
- Freedman, M. & Morris, R. 2002. Tarzan: A Peer-to-Peer Anonymizing Network Layer. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*. pp. 193-206.
- Foster, I. & Iamnitchi, A. 2003. On Death, Taxes, and the Convergence of Peer-to-Peer and Grid Computing. *2nd International Workshop on Peer-to-Peer Systems*.
- Gal, T., Singer, H. & Popkin, L. 2003. The IP War: Apocalypse or Revolution? In *Proceedings of the 2003 ACM Workshop on Digital Rights Management*, pp. 39-46.
- Garnett, N. 2001. Digital Rights Management, Copyright, and Napster. *ACM SIGecom Exchange*. vol. 2, #1, pp. 1-5.
- Giesler, M., Pohlmann, M. 2003a. The Social Form of Napster: Cultivating the Paradox of Consumer Emancipation. *Advances in Consumer Research*. vol. 30, #1, pp. 94-101.

- Giesler, M., Pohlmann, M. 2003b. The Anthropology of File Sharing: Consuming Napster as a Gift. *Advances in Consumer Research*, vol. 30, #1, pp. 273-280.
- Golle, P., Leyton-Brown, K. & Mironov, I. 2001. Incentives for Sharing in Peer-to-Peer Networks. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*. pp. 264-267.
- Good, N. & Krekelberg, A. 2003. Usability and Privacy: a Study of Kazaa P2P File-Sharing. In *Proceedings of the Conference on Human Factors in Computing Systems*, pp. 137-144.
- Gordon, W. & Watt, R. 2003. *The Economics of Copyright: Developments in Research and Analysis*. Edward Elgar Publishing.
- Graziano, M. 2001. The Music Downloading Deluge. *Pew Internet Tracking Report*.
- Greene, S. 2001. Reconciling Napster with the SONY Decision and Recent Amendments to Copyright Law. *American Business Law Journal*.
- Gummadi, P., Saroiu, S. & Gribble, S. 2002. A Measurement Study of Napster and Gnutella as Examples of Peer-to-Peer File Sharing Systems. *Computer Communication Review*. January, vol. 32, #1, p. 82.
- Günnewig, D., Becker, E., Buhse, W. and Rump, N. (Eds.) 2003. *Digital Rights Management - Technological, Economic, Legal and Political Aspects in the European Union*. Springer-Verlag, New York.
- Gupta, M., Judge, P. & Ammar, M. 2003. A Reputation System for P2P Networks. In *Proceedings of the 13th International Workshop on Network and Operating System Support for Digital Audio and Video*. pp. 144-152.
- Habib, A., Xu, D., Atallah, M. & Bhargava, B. 2002 [Verifying Data Integrity in Peer-to-Peer Video Streaming](#). *CERIAS TR 2002-39*, Purdue University.
- Habib, A. & Chuang, J. 2004. Service-Differentiated Peer Selection: An Incentive Mechanism for Peer-to-Peer Media Streaming. *Technical Report, SIMS, UC Berkeley TR-2004-2-HC*.
- Hefeeda, M., Habib, A. & Bhargava, B. 2002. Cost-Profit Analysis of a Peer-to-Peer Media Streaming Architecture, *CERIAS TR 2002-37*, Purdue University, October 2002.
- Hefeeda, M., Afeche, P. & Bhargava, B. 2003. Economics of a Collaborative Peer-to-Peer Infrastructure for Content Distribution, *CS-TR 03-015*, Purdue University, May 2003.
- Hefeeda, M., Habib, A., Botev, B., Xu, D. & Bhargava, B. 2003. PROMISE: Peer-to-Peer Media Streaming using CollectCast. *ACM Multimedia '03*, pp. 45-54, Nov. 2003.

- Heverly, R. A. 2003. The Information Semicommons. *Berkeley Technology Law Journal*, vol. 18, #4, pp. 1127-1190.
- Horne, B., Pinkas, B. & Sander, T. 2001. Escrow services and incentives in peer-to-peer networks. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*. pp. 85-94.
- Hosein, I., Tsiavos, P. & Whitley, E. 2003. Regulating Architecture and Architectures of Regulation: Contributions from Information Systems. *International Review of Law, Computers and Technology*. March, vol. 17, #1, pp. 85-98.
- Hughes, J. & Lang, K. 2004. Issues in Online Focus Groups: Lessons Learned from an Empirical Study of Peer-to-Peer Filesharing System Users, forthcoming in *Electronic Journal of Business Research Methods*.
- Izal, M., Biersack, E. W., Felber, P. A., Urvoy-Keller, G., Al Hamra, A., Garces-Erice, L. 2004. Dissecting BitTorrent: Five Months in a Torrent's Lifetime, *5th Passive and Active Measurement Workshop (PAM2004)*.
- Jensen, R. & Xiao, J. 2001. Customized Financial Reporting, Networked Databases, and Distributed File Sharing. *Accounting Horizons*. September, vol. 15, #1, pp. 209-222.
- Kamvar, S., Schlosser, M. & Garcia-Molina, H. 2003. The Eigentrust Algorithm for Reputation Management in P2P Networks. In *Proceedings of the 12th International Conference on World Wide Web*. pp. 640-651.
- Kant, K. & Mohapatra, P. 2001. Current Research Trends in Internet Servers. *ACM SIGMETRICS Performance Evaluation Review*. September, vol. 29, #2, pp. 5-7.
- Kapur, A.K., Gautam, N., Brooks, R.R., and Rai, S. 2003. Performance and Design of P2P Networks for Efficient File Sharing. In *Proceedings of Industrial Engineering Research Conference*.
- Klampanos, I & Jose, J. 2004. An Architecture for Information Retrieval Over Semi-Collaborating Peer-to-Peer Networks. In *Proceedings of the 2004 ACM Symposium on Applied Computing*. pp. 1078-1083.
- Klein, B., Lerner, A. & Murphy, K. 2002. The Economics of Copyright 'Fair Use' in a Networked World. *American Economic Review*. May, vol. 92, #2, pp. 205-208.
- Koubarakis, M., Tryfonopoulos, C., Idreos, S. & Drougas, Y. 2003. Selective Information Dissemination in P2P Networks: Problems and Solutions. *ACM SIGMOD* vol. 32, #3, pp. 71-76.
- Krishnamurthy, B., Wang, J. & Xie, Y. 2001. Early Measurements of a Cluster-based Architecture for P2P Systems. In *Proceedings of the First ACM SIGCOMM Workshop on Internet Measurement*.

- Krishnan, R., Smith, M., Tang, Z. & Telang, R. 2004. The Impact of Free-Riding on Peer-to-Peer Networks. In *Proceedings of the 37th Hawaii International Conference on System Sciences*.
- Krishnan, S. & Uhlmann, J. 2004. The Design of an Anonymous File-Sharing System Based on Group Anonymity. *Information and Software Technology*, vol. 46, #4, pp. 273-279.
- Kubiatowicz, J. 2003. Extracting Guarantees From Chaos. *Communications of the ACM*. February, vol. 46, #2, pp. 33-38.
- Kwok, S., Lang, K. & Tam, K. 2002a. Peer-to-Peer Technology Business and Service Models: Risks and Opportunities. *Electronic Markets*. vol. 12, #3, pp. 1-9.
- Kwok, S. 2002b. Digital Rights Management for the Online Music Business. *ACM SIGecom Exchanges*. June, vol. 3, #3.
- Kwok, S., Lui, S., 2002c. A License Management Model for Peer-to-Peer Music Sharing. *International Journal of Information Technology & Decision Making*, September, vol. 1, #3, pp. 541-558.
- Lai, K., Feldman, M., Stoica, I. & Chuang, J. 2003. Incentives for Cooperation in Peer-to-Peer-Networks, *Workshop on Economics of Peer-to-Peer Systems*, Berkeley, California, June 2003.
- Lakshminarayanan, K. & Padmanabhan, V. 2003. Some Findings on the Network Performance of Broadband Hosts. In *Proceedings of the 2003 ACM SIGCOMM Conference on Internet Measurement*. pp. 45-50.
- Lam, C. & Tan, B. 2001. The Internet is Changing the Music Industry. *Communications of the ACM*, August, vol. 44, #8, pp. 62-68.
- Landau, M. 2002. Digital Downloads, Access Codes, and US Copyright Law. *International Review of Law, Computers & Technology*. July, vol. 16, #2, pp. 149-171.
- Lechner, U. & Hummel, J. 2002. Business Models and System Architectures of Virtual Communities: From a Sociological Phenomenon to Peer-to-Peer Architectures. *International Journal of Electronic Commerce*, vol. 6, #3, pp.41-53.
- Lee, J. 2003. An End-User Perspective on File-Sharing Systems. *Communications of the ACM*. February, vol. 46, #2, pp. 49-54.
- Lee, Y., Oh, C. & Park, E. 2002. Intelligent knowledge discovery in Peer-to-Peer File Sharing. In *Proceedings of the 11th International Conference on Information and Knowledge Management*. pp. 308-315.
- Leiner, B., Cerf, V., Clark, D., Kahn, R., Kleinrock, L., Lynch, D., Postel, J., Roberts, L. & Wolff, S. 1997. The Past and Future History of the Internet. *Communications of the ACM*, vol. 40, #2. pp. 102-108.

- Lessig, L. 2001. *The Future of Ideas: The Fate of the Commons in a Connected World*. Random House, New York.
- Lessig, L. 2003. *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. Penguin Books.
- Litman, J. 2001. *Digital Copyright: Protecting Intellectual Property on the Internet*. Prometheus Books.
- Litman, Jessica D., 2003. "Sharing and Stealing" (November 23, 2003). <http://ssrn.com/abstract=472141>
- Liu, S. & Kwok, S. 2002. Interoperability of Peer-to-Peer File Sharing Protocols. *ACM SIGecom Exchanges*, August, vol. 3, #3, pp. 25-33.
- Loo, A. 2003. The Future of Peer-to-Peer Computing. *Communications of the ACM*. September, vol. 46, #9, pp. 56-61.
- Lu, J. & Callan, J. 2003. Content-Based Retrieval in Hybrid Peer-to-Peer Networks. In *Proceedings of the 12th International Conference on Information and Knowledge Management*. pp. 199-206.
- Lv, Q., Cao, P., Cohen, E., Li, K. & Shenker, S. 2002a. Search and replication in unstructured peer-to-peer networks. In *Proceedings of the 16th international conference on Supercomputing*. pp. 84-95.
- Lv, Q., Ratnasamy, S., & Shenker, S. 2002b. Can Heterogeneity Make Gnutella Scalable. In *Proceedings of the 1st International Workshop on Peer-to-Peer Systems, (IPTPS '02)*. (Cambridge, MA, March 2002).
- Ma, R., Lee, C., Lui, J. & Yau, D. 2003. Incentive P2P networks: a protocol to encourage information sharing and contribution. *ACM SIGMETRICS Performance Evaluation Review*. September, vol. 31, #2, pp. 23-25.
- Mackenzie, E. & Goldman, K. 2000. Computer abuse, information technologies and judicial affairs. In *Proceedings of the 28th annual ACM SIGUCCS conference on User services: Building the future*. pp. 170-176.
- Mulligan, D., Han, J. & Burstein, A. 2003. How DRM-Based Content Delivery Systems Disrupt Expectations of "Personal Use". In *Proceedings of the 2003 ACM Workshop on Digital Rights Management*. pp. 77-89.
- Nejdl, W., Wolpers, M., Siberski, W., Schmitz, C., Schlosser, M., Brunkhört, I. & Loser, A. 2003. SuperPeer-Based Routing and Clustering Strategies for RDF-Based P2P Networks. *12th International Conference on World Wide Web*. pp. 536-543.

Netanel, N. 2003. Impose a Noncommercial Use Levy to Allow Free P2P File-Swapping and Remixing. *Harvard Journal of Law & Technology*, Fall 2003, 17.

Oram, A. (Ed.) 2001. *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*. O'Reilly & Associates; 1st edition (March 15, 2001).

Orlikowski, W. & Iacono, C. 2001. Desperately Seeking the "IT" in IT Research - A Call to Theorizing the IT Artifact. *Information Systems Research*, June, vol. 12, #2, pp. 121-134.

Park, J. & Hwang, J. 2003. Role-based access control for collaborative enterprise in peer-to-peer computing environments. In *Proceedings of the eighth ACM Symposium on Access Control Models and Technologies*. pp. 93-99.

Peitz, M. & Waelbroeck, P. 2003. Piracy of Digital Products: A Critical Review of the Economics Literature. November. *CESifo Working Paper Series No. 1071*.  
<http://ssrn.com/abstract=466063>

Portmann, M. & Seneviratne, A. 2003. Cost-effective broadcast for fully decentralized peer-to-peer networks. *Computer Communications*, July, vol. 26, #11, pp. 1159-1168.

Ranganathan, K., Ripeanu, M., Sarin, A. & Foster, I. 2003. *To Share or Not to Share: An Analysis of Incentives to Contribute in Collaborative File Sharing Environments*. Workshop on Economics of P2P Systems, Berkeley, California, June 2003.

Ratnasamy, S., Francis, P., Handley, M., Karp, R. & Schenker, S. 2001. A Scalable Content-Addressable Network. In *Proceedings of 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*. pp. 161-172.

Ripeanu, M. Foster, I. & Iamnitchi, A. 2002. Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design. *IEEE Internet Computing Special Issue on Peer-to-Peer Networking*. Vol. 6, #1, pp. 50-57.

Robinson, L. & Halle, D. 2002. Digitization, the Internet, and the Arts: eBay, Napster, SAG, and e-Books. *Qualitative Sociology*. September, vol. 25 #3, pp. 359-383.

Samuelson, P. 2003. DRM (and, or, vs. ) the Law. *Communications of the ACM*. April, vol. 46, #4, pp. 41-45.

Samuelson, P., 2004. What's at Stake in MGM v. Grokster? *Communications of the ACM*, February, vol. 47, #2, pp. 15-21.

Saroiu, S., Gummadi, K., Dunn, R., Gribble, S. & Levy, H. 2002. An Analysis of Internet Content Delivery Systems. In *Proceedings of the 5th Symposium on Operating Systems Design and Implementation*. vol. 36, issue SI, pp. 315-327.

Saroiu, S., Gummadi, K. & Gribble, S. 2003. Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts. *Multimedia Systems*. August, vol. 9, #2, pp. 170-184.

Schoder, D. & Fischbach, K. 2003. Peer-to-Peer Prospects. *Communications of the ACM*. February, vol. 46, #2, pp. 27-30.

Serjantov, A. 2002. Anonymizing Censorship-Resistant Systems. In *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02)*, MIT Faculty Club, Cambridge, MA, USA, March 2002.

Shneidman, J. & Parkes, D. 2003. Rationality and self-interest in peer-to-peer networks. In *Proceedings of 2nd International Workshop on Peer-to-Peer Systems*.

Snir, E. 2003. The Record Industry in an Era of File Sharing: Lessons from Vertical Differentiation. In *Proceedings of the 24th International Conference on Information Systems (ICIS)*. pp. 72-84.

Sripanidkulchai, K., Maggs, B. & Zhang, H. 2002. Enabling efficient content location and retrieval in peer-to-peer systems by exploiting locality in interests. *ACM SIGCOMM Computer Communication Review*, vol. 32, #1.

Stallman, R. 2002. *Free Software, Free Society: Selected Essays of Richard M. Stallman*. Free Software Foundation.

Stoica, I., Morris, R., Liben-Nowell, D., Karger, D., Kaashoek, M., Dabek, F. & Balakrishnan, H. 2003. Chord: a scalable peer-to-peer lookup protocol for internet applications. *IEEE/ACM Transactions on Networking*. February, vol. 11, #1, pp. 17-32.

Strahilevitz, L. 2003. Charismatic Code, Social Norms, and the Emergence of Cooperation on the File-Swapping Networks. *Virginia Law Review*, vol. 89. 505.

Swanson, E. & Ramiller, N. 1993. Information Systems Research Thematics: Submissions to a New Journal, 1987-1992. *Information Systems Research*. vol. 4, #4, pp. 299-330.

Tanaka, H. 2001. Post-Napster: Peer-To-Peer File Sharing Systems: Current And Future Issues On Secondary Liability Under Copyright Laws In The United States And Japan. *Loyola of Los Angeles Entertainment Law Review*, vol. 37, # 49.

Triantafillou, P., Xiruhaki, C. & and Koubarakis, M. 2002. Efficient Massive Sharing of Content among Peers. *IEEE Workshop on Resource Sharing in Massively Distributed Systems*, (RESH '02), July 2, 2002, Vienna, Austria.

Turner, D. & Ross, K. 2004. A Lightweight Currency Paradigm for the P2P Resource Market. *Seventh International Conference on Electronic Commerce Research*.

Vaucher, J., Babin, G., Kropf, P. & Jouve, T. 2002. Experimenting with Gnutella Communities.

In *Conference on Distributed Communities on the Web*. (DCW 2002), Sydney, Australia, 2002, pp. 85-99.

Vishnumurthy, V., Chandrakumar, S. & Sirer, E. 2003. KARMA: A Secure Economic Framework for P2P Resource Sharing. In *Proceedings of the Workshop on the Economics of Peer-to-Peer Systems*, Berkeley, California, June 2003.

Whinston, A., Parameswaran, M. & Susarla, A. 2001. P2P Networking: An Information Sharing Alternative. *IEEE Computer*. July, pp. 1-8.

Yang, B. & Garcia-Molina, H. 2002. Efficient Search in Peer-to-Peer Networks. In *Proceedings of the 22nd International Conference on Distributed Computing Systems*, Berkeley, California, June 2003.

Yang, C. 2003a. Peer-to-Peer Architecture for Content-Based Music Retrieval on Acoustic Data. In *Proceedings of the 12th International Conference on World Wide Web*. pp. 376-383.

Yang, B., Kumvar, S. & Garcia-Molina, H. 2003b. Addressing the Non-Cooperation Problem in Competitive P2P Systems. In *Workshop on Economics of Peer-to-Peer Systems*, Berkeley, California, June 2003.

## **Appendix - Database Naming and Numbering Conventions**

**PaperID:** first four characters of lead author's name, concatenated with year of publication.

Multiple papers by one author in a given year have an alphabetic character appended.

Example: Giesler's two papers of 2003 are Gies2003a and Gies2003b. In the event that two different papers by different authors with similar or the same last names in the same year coincidentally would have the same code, the alphabetic character is included to distinguish them.

**ReferenceID:** same format as PaperID

**PublisherID:** acronym of title. Example: CACM

Conference acronyms attempt to conform to the conference's own usage

**TypeID:** 3 numeric characters. 1st character is same as Framework number.

Last two characters just number the types within that framework.

**AuthorID:** is full last name, concatenated with first letter of first name