

TOOLS OF CHANGE FOR PUBLISHING
RESEARCH REPORT:

Impact of P2P and Free Distribution on Book Sales

Impact of P2P and Free Distribution on Book Sales

by Brian O'Leary

Copyright © 2009 Brian O'Leary, All rights reserved.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (<http://my.safaribooksonline.com/>). For more information, contact our corporate/institutional sales department: 800-998-9938 or corporate@oreilly.com.

Editor: Mac Slocum

Interior Designer: Mark Paglietti

Illustrator: Robert Romano

ISBN: 9780596157876

1. Challenging Notions of "Free"

Book publishers have long used free content as part of their marketing and selling efforts, with the vast majority of free content distributed in printed form. Almost every publisher has distributed galleys, advance reading copies, blads and sample chapters to help promote book sales, using the size of the press run to control the extent to which content is given away.

As the use and effectiveness of Web-based search has grown, publishers have increasingly given various sites permission to offer digital content, typically sample chapters, segments or even entire books with limits on how much content can be freely accessed. Online retailers like [Amazon](#) and [Barnes & Noble](#) offer product-specific features ("Search Inside") to let customers sample content before buying.

Few publishers have embraced free distribution of an entire book, fearing that doing so will cannibalize sales of print books. In some markets where book piracy is more widely an issue (college textbooks provide an example), publishers have sought to block unauthorized distribution, paralleling an anti-piracy approach seen in the music industry.

Because participants and observers have raised many questions about the impact of free distribution on paid content, we undertook this research effort to challenge established or assumed notions of the impact of free distribution on book sales. Taking up the research now is important for three reasons:

1. The sophistication of ebook readers is increasing, and as the devices improve, they are bringing more content into the marketplace. While a relatively

small mix of titles has been made available digitally to this point, creation and consumption of digital content is increasingly a fact of life. Improving our understanding of the value or cost of freely available content can help steer publishers toward the right answers as the amount of such content grows.

2. The proliferation of digital content is aided by a commitment by search engines and online retailers who see interest in free distribution of at least partial content among their online audiences. Establishing a baseline impact of free content on visibility, discoverability and ultimately sales may also help inform publishers' thinking about their marketing mix.
3. There is significant discussion about the extent to which free online content costs publishers sales they otherwise would have had. Here, we establish a set of guidelines that can continue to inform this discussion, providing a vehicle to confirm, refute or extend the findings presented here.

Designing credible experiments whose results can be analyzed and compared depends on the ability to test enough titles to control for a range of variables. This project began with an assessment of work already done, most notably by O'Reilly Media and Random House. That assessment led to an analysis of the attributes that might be captured and compared when planning and conducting experiments with the distribution of free content.

1.1. Project Structure

This research and analysis is expected to test and perhaps change perceptions about free distribution of digital content. To help that happen, we wanted to develop and use a consistent, documented approach. While the value of free distribution and the potential impact of peer-to-peer (P2P) file sharing will continue to be a matter of some discussion and debate, the tests described here provide a basis for informed review as well as continued refinement using an approach that:

- Documents, characterizes and assesses work already done to promote book sales through the use of freely distributed content.
- Addresses data quality as an underlying issue in the ongoing debate about the impact of free content.
- Analyzes, documents and shares the results of a set of experiments involving free distribution of content.
- Provides a sense of the implications and next steps that can help advance publishers' understanding of when and how free distribution may make sense for them.

This research document reflects the project methodology, which is data-driven and intended to share knowledge freely and grow the number of publicly reported experiments with free distribution. Because publishers are just starting to feel the impact of file sharing, we took an approach that identified ways publishers could steer a profitable course in an environment that increasingly relies on free content as part of the overall marketing and sales mix.

For publishers not familiar with P2P distribution, we also documented and characterized the nature of the P2P universe. While much of the work done to date involves deliberate release of PDFs or ebook formats, the role of pirated content may grow over time.

It will be useful and instructive to broaden the sample set by recruiting other trade publishers to accompany work beyond O'Reilly and Random House. The results to date include assessments of titles published

by both houses, but data gaps remain that can be addressed by including a wider range of titles and tests.

1.2. Findings and Recommendations

With respect to the impact of free and pirated content, we are able to offer five clarifying observations:

- We propose a less binary model to evaluate the use of free: "white" – "gray" – "back channel," with differing levels of risk and benefit for publishers looking to grow revenue and build an author's reputation.
- Measures must evolve and expand to include the impact on hard-copy sales, changes in digital sales, conversion from trialware copies and perhaps other measures of awareness, engagement and referrals.
- With respect to the impact of free distribution, the experience of books does not appear to directly parallel other media (music, movies, others). The primary difference is the engagement required when reading a book. Digital versions of books demand a different interface, something that may be a barrier to downloading.
- The near-term threat posed by P2P file distribution may not be as significant as publishers have assumed. Our research indicates that books make up a very small share of the content available on P2P sites, and there can be a significant lag (weeks to months) before the content of many pirated books makes its way to P2P sites.
- As well, the degree of technological sophistication required to find and download a book using P2P technology is still significantly greater than most Internet users possess. BitTorrent, the most common P2P protocol, is embedded only in one relatively obscure Web browser (Opera). This also suggests that the range of books that might be threatened by P2P networks is limited in the near term.

In addition, our methodology helped us uncover the following:

- Tests of free content are varied, but they can be compared, if you adopt a sufficiently robust test matrix.
- What matters most has yet to be established, but books appear to be most readily characterized by: markets served; the timing, type of book, degree of author involvement, method of free distribution and the nature of free content (protected or not).
- What works also depends on the characteristics of the content. This research study uncovered at least three ways in which content distribution might be grouped: audience-specific; subject-specific; and author-specific. Further research may refine these subsets.
- Tests provide useful ways to validate (or refute) hypotheses, several of which are laid out in [Section 1.4, "Segmenting Different Types of Book Marketing"](#).

As work on this project continues, we expect that this list will both evolve and grow.

1.3. Characterizing and Assessing the Use of Free Content

In addition to the free content (galleys, advance reading copies and the like) that almost all publishers give away to promote titles, a range of books have also been promoted using digital content. Understandably, these experiments have been conducted in a variety of ways, and documentation of various tests varies in depth and completeness. In some cases, limited documentation has weakened arguments on both sides of the debate.

In marketing their titles, book publishers consistently face an abundance of content and a variety of channels through which customers can purchase books. To overcome these challenges, publishers have used their marketing efforts to grow discovery (the ability to learn about a title) and access (the ability to buy the title at a time and place of a customer's choosing).

Some marketing efforts are particularly effective at promoting discovery (see [Figure 1, "Book marketing:](#)

[growing content discovery and access"](#)). For example, an author who appears on Oprah Winfrey's show is likely to gain significant exposure and build awareness for a given title. That appearance, however, doesn't immediately link to sales: the viewer has to do something else to buy the book.

Other marketing efforts can help build access. An author tour that includes book signings in individual stores may have an immediate effect on sales, but the nature of the promotion means that a publisher reaches only one store at a time (low discovery). While word of mouth may build, promoting discovery, the impact on access is limited to the period surrounding the author's appearance.

The most effective marketing investments advance both discovery and access simultaneously. Selling art books in a museum store, using Amazon to promote a title with both content and a special offer, or sponsoring in-store cooperative marketing displays, all provide opportunities to (with the right offers) increase discovery and access at the same time.

In this framework, digital distribution of content offers the opportunity to increase awareness and trial but does not necessarily provide access. In that regard, it may be seen as an opportunity to use content as a lower-cost marketing option. The impact of free content on book sales then becomes the open question.

Our initial sample measured the sales of eight trade titles published across a total of 12 formats (variously, hard cover, mass market, compact disc, trade paperback and audiobook download). Our analysis tracked sales in the four weeks prior to each promotion and compared it to sales during the promotional period (typically one week, though longer in some cases) and the four weeks after a promotion ended.

We found that free distribution, on average, coincided with sales growth of 19.1 percent in the promotional period and 6.5 percent across the combined promotional period and the following four weeks. The combined promotional and post-promotional period was typically a total of five weeks; in three tests it was a total of seven weeks.

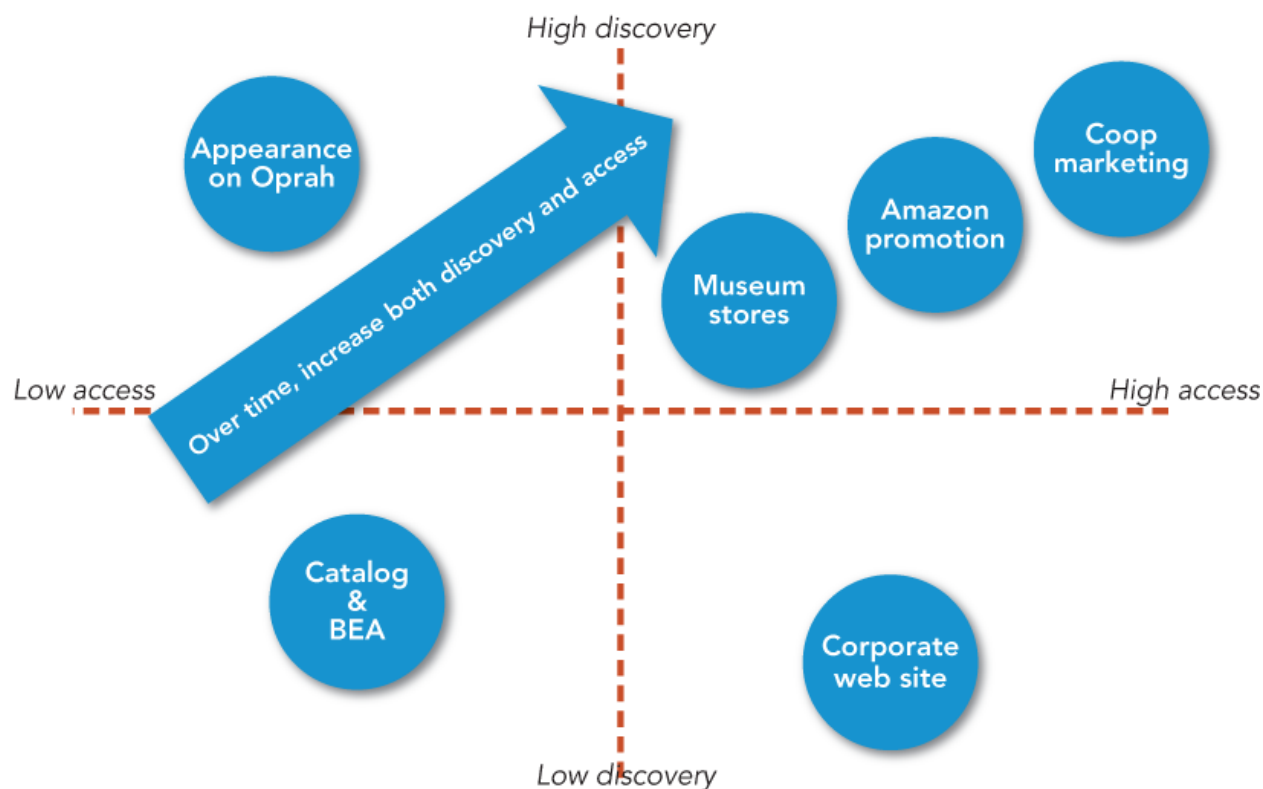


Figure 1. Book marketing: growing content discovery and access

Within this sample, results varied widely. Across the promotional (one to three weeks) and subsequent four-week post-promotional period, results varied from a sales increase of 155 percent to a decline of 74 percent for a first-time novelist. The next section presents title-specific results.

1.4. Segmenting Different Types of Book Marketing

As noted, analyzing the work done by various publishers prior to this research study uncovered at least three ways in which content distribution might be grouped: audience-specific (e.g., science fiction fans); subject-specific (e.g., gardening, or rewiring an older home); and author-specific (see Figure 2, “Options to focus marketing”). Further research may uncover or help define other sets or subsets of these three group-

ings, but these appear to be the right groupings for the available test data.

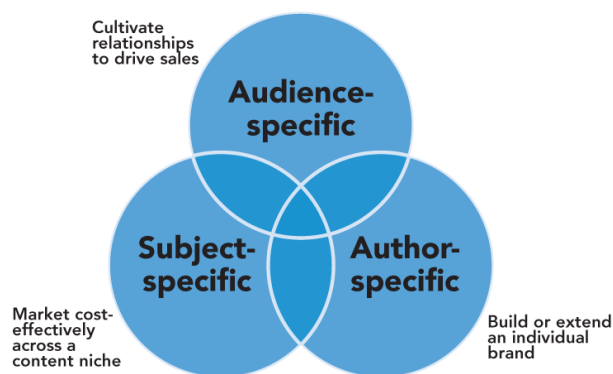


Figure 2. Options to focus marketing

1.4.1. Examples of audience-specific promotions

Audience-specific marketing allows authors, publishers or a combination of the two to cultivate relationships with potential buyers. Leveraging content free of charge, whether in whole or in part, can be an effective way to market within an informed or topic-driven audience. It can also help potential buyers evaluate whether a particular title offers the desired content. In the case of series science fiction, for example, distributing content for the first book in a series free of charge may spur interest in purchasing other books in the line.

For four days in March 2008, a Random House imprint offered a PDF download of a science fiction/horror title, a promotion that preceded the book's publication date by a few days. The download was hosted at Random House and also offered by Amazon, Barnes & Noble and Sony. The title was promoted heavily around the time of the download and sales grew by a third at the time of the free download before dropping to lower, but not small numbers, in the weeks that followed.

Over a two-week period in early May 2008, another Random House imprint offered free downloads of several ebook formats for a science fiction title with a film tie-in. The promotion coincided with the on-sale date of the title's ebook format. In the promotional week, the offer coincided with a 4 percent increase in sales of the mass-market paperback edition that had been released a year earlier. In the four weeks that followed, sales of the print edition were more than 40 percent higher. Sales of the ebook were twice what they had been before the promotion was announced.

In the prior section we noted that leveraging content free of charge, whether in whole or in part, can provide an effective way to market within an informed or topic-driven audience. These examples support that claim. The science fiction/horror author is known within a community of science-fiction enthusiasts, and the download appeared to contribute to a fast launch of a book that appealed to the community. In a similar way, releasing a free copy of a science fiction ebook may have spurred renewed interest in the mass-market edition. In each of these cases, the hypothesis

would suggest that free distribution encouraged sampling and ultimately sale of the print edition.

1.4.2. Examples of subject-specific promotions

Subject-specific marketing provides publishers with an opportunity to more cost-effectively market across a specific content niche. Here, free content establishes credibility and spurs interest in the publisher as brand. This brand awareness then leads to trial and loyalty. For example, O'Reilly employs topic focus and free distribution to develop an audience that values its content enough to purchase print and digital products.

In April 2008, a Random House imprint offered a full PDF download of a first novel about a spiritual quest. The free version was made available on a Random House Web site from the date of publication, and the download was promoted on a variety of Web sites. During the evaluation period, the title significantly improved its sales standing on Amazon and appeared to sell most effectively through this account. Overall sales dipped somewhat (14 percent) during the download period but grew in the subsequent four weeks to show a slight gain (1 percent) in average sales.

In a similar vein, another Random House imprint provided readers with an opportunity to download a free copy of a title offering financial advice with an ethical emphasis. The free file was given away for a day in April 2008, two weeks after the title's on-sale date. After the download, which was also promoted by Amazon and Sony, sales of the print and CD-format audiobook were 29 percent higher, stemming a prior decline in print sales.

In 2007, Steven Poole released a free PDF of his 2000 book (updated and expanded in 2004), *Trigger Happy*, a history of video gaming. Anyone downloading the PDF had a chance to donate an amount of their choosing. While very few people elected to donate, Poole concluded that "downloads don't cannibalize print sales; if anything, they encourage them." He also offered a caution: "Giving away your work in the same format in which you hope to sell it is a dangerous game, if that's how you hope to make a living."

Poole's beliefs about the extent to which digital content can grow sales of otherwise obscure titles is supported by a 2006 analysis conducted by O'Reilly. When O'Reilly ranked the titles it sold by decile, it found that 7 percent of the page views for its Safari Books Online subscription service involve titles that are not selling at all in print; 20 percent of access involves books that generate only 5 percent of print book sales. The conclusion: digital distribution of at least some types of content offers publishers the opportunity to generate interest in, and promote sales of, titles that may have passed the point at which **it made sense to keep them in print**.

In a 2008 interview with *Tools of Change for Publishing*, Peter Kent, senior vice president for U.S. operations at DNAML, a supplier of ebook software, underscored the impact of subject-specific marketing. "It's no longer a gentleman's game in which everyone hands over their books to a bookstore, and then they all compete on the same level," he noted. "In the future the more aggressive publishers are going to go out and find book buyers even before the buyers have thought about buying!"

At least two hypotheses may be constructed based on these examples. The first suggests that subject-specific titles may sell better with digital promotions, including the use of free content or downloads, because the target is more likely to turn to the Internet to obtain information. In the second hypothesis, the use of digital content may spur the sale of print and other physical products through online merchants, as the audience will look to buy in the medium that most closely resembles the one in which it found the content.

1.4.3. Examples of author-specific promotions

Author-specific marketing provides authors with the potential to develop and exploit content to build an individual brand. If obscurity is the primary obstacle to the success of a new release, an author just starting out may be willing to share content freely to promote awareness. Over time, that awareness and trial may reduce the need for continued use of free content.

A print analogue applies: advance reading copies may provide a new author with an outsized benefit when it comes to visibility, while the same promotion for an established author may have only minimal impact on overall awareness. For this reason, the release of freely distributed material or the allowance of pirated content for an author who is widely established may be seen differently from that affecting a new or unknown author, even at an established imprint.

In February 2008, Random House offered free downloads of a personal finance best seller. The book had been on sale for a year and had sold well to that point. The free download was publicized widely and coincided with an immediate tripling of print sales. During the measured period, overall print sales were up 155 percent.

Also in February 2008, Random House offered a free download of a high-profile literary first novel. The book had been published for six weeks at the time of the download. For three days the PDF was offered freely on the author's Web site as well as the Web sites for Amazon, Barnes & Noble and Books-A-Million. The free download coincided with an increase in sales of the audiobook (available in CD format) and an ebook version of the title. Hardcover print sales, which had been declining, did continue to decline after the promotion. The data does not clearly establish whether the free download had any impact (positive or negative) on hardcover book sales.

Starting in March 2008, Crown/Harmony offered an 80-page PDF that included excerpts from a new-age religious best seller as well as new content related to the book. Content was featured on sites hosted by the author and on an author-branded site. During the promotion, a decline in print sales slowed. Results may have been affected by other promotions for the same title at this time.

O'Reilly author David Pogue, while not part of a formal experiment, **has joined the ongoing debate** about the value of free with a concern that unpaid access to his titles (through pirated content) is costing him print-book sales. His concern, which has been challenged by others, may reflect two competing interests for author-specific marketing: those authors

with significant existing platforms may not need or want the promotion available through the use of free content; and persistent availability of free content may supplant sales when the author is already branded.

These cases point to the role that a significant platform or celebrity status can play in assessing the impact on sales following the offer of free content.

1.5. Impact of Prior Work on Design of This Research Study

In collating the results of these prior experiments, it became clear that while the various tests differed widely from one another on a variety of attributes, they also consistently shared attributes that could be used to better understand cause and effect. The science fiction/horror PDF and the science fiction ebook distribution differed in scale, the period for which free content was made available and the relative intensity of promotion. However, they both appealed to an audience that knew what it could expect from the author or series. Other content genres (romance novels, for example) might be expected to show similar behaviors.

Similarly, the time periods for releasing free content for the first-time novel about a spiritual quest and the title offering financial advice with an ethical emphasis varied significantly, but they share subject-specific characteristics that may have trumped the differences in the tests. As other tests are planned, certain titles may be effectively marketed using free content when the person searching for subject matter isn't necessarily looking for a book. The second lesson may be that straightforward sales access inside the relevant medium (in this case, the Internet) becomes more important as publishers pursue incremental sales.

While there is a tendency to dismiss the success of promotions driven by celebrity authors as the exception, the opportunity to leverage authors' reputation may extend beyond the small cadre of well-known individuals. The ability to capitalize on a cross-media platform, even in a limited field, may be enough (David Pogue's recognition as a book, newspaper and Web columnist applies). As well, it may be possible to leverage major accounts in a coordinated fashion to

drive awareness and create a sense of "author ubiquity." Finally, a small amount of "something new," long used in marketing books, can apply here, as well.

As these prior tests were reviewed, it became increasingly clear that free distribution certainly is "**more complicated than you think.**" To address that apparent complexity and promote understanding of what elements divergent tests may share, we began to develop the test matrix shown as in Figure 3, "**Free content matrix**". Beyond tracking the three marketing attributes (audience, subject and author) discussed here, the matrix also considers market subsets, types of distribution, file characteristics and digital file formats. While these elements can be expected to affect the value of free distribution in a variety of ways, tracking each experiment in as much detail as possible allows downstream analysis of a mix of tests with a variety of comparable attributes (the impact of free distribution on front-list vs. back-list titles, for example).

The matrix approach also provides a way to track and incorporate future experiments, growing the breadth and depth of the sample set in areas that have not been tested, or for which new experiments can add or challenge the validity of earlier findings. Ultimately, this provides a foundation for continued experimentation and informed collaboration as publishers and authors work to find the right mix of paid and free content.

Overall, research and experiments done to date have been helpful in establishing that some titles may benefit from freely distributed digital content. As we learned of a range of "free" experiments, we found that different methods and measures used have made it difficult for publishers to compare results.

Without comparability, the sample sizes for any given type of free distribution are often limited. Because the lack of data and the inflammatory nature of the topic have often led to strong opinion, we sought to evaluate the data we do have and ask "*what does the use of free content mean for book sales?*"

1.6. Analysis of P2P Impact

Many of the tools used for free distribution of content (PDF downloads, trialware, watermarked audio files

Title	Market			Timing		Type		Involvement		Distribution			File characteristics			Content characteristics			Format	
	Adult	YA	Children	Frontlist	Back list	Fiction	Non-fiction	Significant platform	Engaged author	Deliberate	Limited	Pirated	Locked (DRM)	Trial-ware	Open	Audience-specific	Subject-specific	Author-specific	PDF or ebook	MP3
(Science fiction/horror)	✓			✓		✓				✓					✓	✓		✓	✓	
(Science fiction/film tie in)	✓	✓			✓	✓		✓		✓					✓	✓			✓	
(Spiritual quest)	✓			✓		✓				✓					✓	✓	✓		✓	
(Fin advice, ethical emphasis)	✓			✓			✓			✓					✓		✓		✓	
(Personal finance best seller)	✓				✓		✓	✓	✓	✓					✓		✓		✓	
(Literary first novel)	✓			✓		✓				✓					✓				✓	
(New age/rel best seller)	✓				✓		✓	✓	✓	✓				✓					✓	
Ajax	✓			✓			✓					✓			✓		✓		✓	
Python	✓			✓			✓					✓			✓		✓		✓	
C Sharp	✓			✓			✓					✓			✓		✓		✓	
Office	✓			✓			✓					✓			✓		✓		✓	
Dojo	✓			✓			✓					✓			✓		✓		✓	
Linux	✓			✓			✓					✓			✓		✓		✓	
Ubuntu	✓			✓			✓					✓			✓		✓		✓	

Figure 3. Free content matrix

and the like) are familiar to publishers. However, the nature and characteristics of P2P file sharing networks are not understood as well. While sharing book content has not yet reached a level at which publishers have focused on these networks or how they work, the volume of book content shared through such sites is expected to grow. In our research, we have also tried to understand the impact of P2P file sharing on the sales of certain titles.

Titles initially published by O'Reilly in 2008 were monitored to determine when they were uploaded to [The Pirate Bay](#), [Demonoid](#), or [Scribd](#). Only eight frontlist titles were uncovered. On average, the initial upload was far from immediate, coming 20 weeks after publication date. Lag times ranged from a low of four weeks to a high of 43 weeks (see Figure 4, "Lag time before seeding varies").

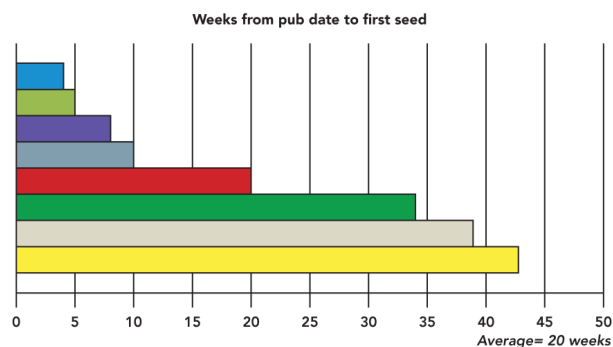


Figure 4. Lag time before seeding varies

Print sales of the eight seeded titles were measured for a total of eight weeks: four weeks before and four weeks after the first seed was located. The week in which a seed was first tracked was not included in either the "before" or "after" periods. On average, sales in the four weeks after the initial seeding were 6.5 percent higher than in the four weeks prior to the

P2P upload. Sales results ranged from an increase of 18.2 percent for a title that had been published for 43 weeks, to a decline of 33.1 percent for a title that had been published 20 weeks earlier.

The data set showed no correlation between pre-seed sales volume and post-seed sales changes. This may reflect the impact of a small sample. There was a stronger correlation (0.74) between the average number of seeds and post-seed print sales growth. Again, causality is not implied.

The volume of upload and download activity at any measured period was generally small. The number of seeds (files made available on a P2P network) ranged from 0 to 43, with an average of 15 seeds. The number of seeds peaked quickly (see Figure 5, “The number of seeds peaked quickly”), typically in the third week after the first seed was uncovered.

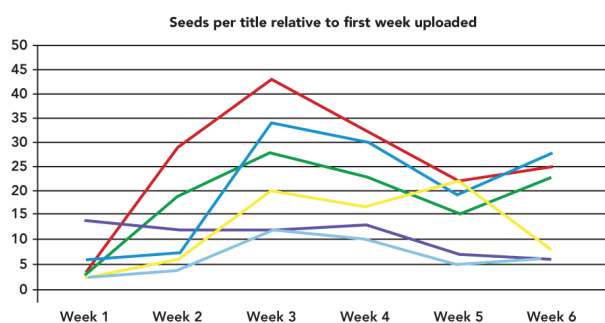


Figure 5. The number of seeds peaked quickly

The number of leeches (active downloads of a file) ranged from 0 to 13, with an average of 2 leeches when monitored. In comparison, the sales for print titles averaged 171 units in the week that seeds first appeared and 154 units per week in the four weeks that followed an initial upload. As with seeds, the number of leeches peaked quickly (see Figure 6, “The number of leeches peaks immediately and quickly declines”), typically in the second week after a seed was first noted.

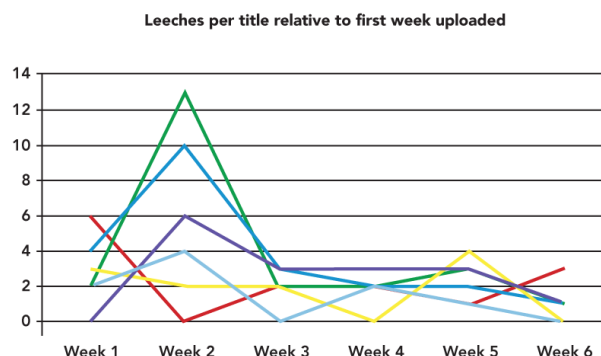


Figure 6. The number of leeches peaks immediately and quickly declines

This data suggests that the presence of files on file-sharing sites may not be uniformly harmful to print sales. The number of titles uploaded at any given time is relatively small. In addition, the volume of both seeds and leeches is low relative to unit sales of print versions of uploaded titles. As the sample size is expanded, it will be possible to draw more detailed analysis between presence on P2P sites and print sales of titles.

1.7. Establishing and Expanding Measures of Success

Developing a model that describes instances in which free distribution works or may not work is the ultimate objective of the research conducted for this project. To do that, we first established expected measures of success, outlined in this section. While the approaches mentioned here are not exhaustive and can be expanded, they provide the basis for evaluating experiments conducted for this study.

1.7.1. Impact on hard-copy sales

Generally, the primary measure of success is strength in print sales, measured in units sold as reported using point-of-sale (POS) data supplied by retailers. To establish a baseline, we gathered data on print sales four weeks before free distribution began, then the sales during the promotion and four weeks after. Weekly sales during each of these periods are averaged to provide a uniform basis for comparison when calculating any gain or loss in print sales.

In some cases, baseline data (four weeks prior to the release of a digital file) is not available. If a shorter baseline is available, this period was used and noted in the research. When no baseline period is available (a digital file is released as the print title goes on sale), the sales data is collected during and after the digital promotion, and the sales profile is analyzed against comparable titles (other books in a series, for example, or other books by the same author).

Because access to digital downloads and pirated content may persist well beyond an established promotional period, the research data may underestimate the gain or loss in print sales following a digital content offer. This is a limitation in the current study that may be addressed over time by collecting ongoing data across a range of titles.

1.7.2. Digital sales

Although the primary measure of success is a positive impact on print sales, prior research and experiments suggests that free content promotions may spur sales of digital content. Where data is available, ebook sales were evaluated in the same way that changes in print sales were assessed. Most titles sell relatively few copies of digital files (ebooks are the primary example), so increases or decreases in sales volume can be somewhat more difficult to validate. Where volumes are small, the statistical volatility is noted.

1.7.3. Conversion from trialware copies

While trialware tests are more common in the software industry, ebooks are becoming book publishing's software option. It is possible to distribute copies of ebooks as complete files whose content can be read in full only after it is unlocked. The value of a trialware promotion can be established by evaluating the rate at which distributed content is purchased. While data developed for this Rough Cut analysis did not include any tests of trialware formats, it is clear that publishers interested in using trialware as an option will need to develop more robust direct-response functions (in-house or outsourced) that take into account the cost of the test and the revenues obtained through conversion.

1.7.4. Other measures

While the first three sets of measures represent "hard" data (print and digital sales; conversion from free content), there are also new or emerging ways of establishing progress or success, especially if the purpose is marketing and branding.

Publishers might want to look at page views, site traffic or the ratio of downloads to page views, among other measures. All of these measure the impact of efforts to grow awareness of or interest in an author or subject, or to help cultivate an audience. This data can be made more useful in context: how many visitors went on to download a file, or did the presence of free content help grow time on a site or the number of page views recorded? These measures are drawn from a somewhat more open-ended list, starting with broader questions about the overall goals of the publisher and author.

Comparably, publishers or authors with a social media focus may want to look at referrals or recommendations. The higher the number and velocity of viral responses, the more likely it is that free content promotions have reached the right audience. Ultimately, we would want these recommendations to convert to sales, an objective that would also be best measured by a more comprehensive approach to direct-response marketing.

Finally, publishers and authors interested in buzz now have tools to measure it more explicitly. Mentions (favorable or unfavorable) across the Web can be tracked using a variety of services, many of them free or low-cost (blog search engines, news alerts, Google Trends, etc.). Some are audience- or subject-specific and can be helpful in building the reputation of a publisher or an author.

Because these measures vary widely and are content-sensitive, they were approached on a case-by-case basis. No attempt was made to establish cause and effect with respect to sales, although over time and with an enhanced data set, making such correlations may become more feasible.

2. Background on Piracy and the Use of Free Content

While some publishers have experimented with free digital distribution of book content, the tests to date are limited and the results have been at times inconclusive or contradictory. In this section, we consider some of that history, the viewpoints of publishers (including O'Reilly), and the rationale for revisiting the topic of free distribution of digital content.

2.1. File Sharing Outside Publishing

Section 3 provides an overview of the evolution of file-sharing from its origins in the 1980s through today. While the roots of file sharing are and remain legitimate, the "many-to-many" nature of Web content sharing opens the door to unauthorized distribution of intellectual property. This potential threat has been felt across several industries whose content is either bought directly (music, subscription cable) or sold against (network television). These industries have experience with unauthorized distribution that predates that of book publishers and may suggest approaches that can, or can't, be used when considering pirated content.

Both the music and motion picture industries have fought to stop or significantly hinder illegal file sharing. The Recording Industry Association of America (RIAA) has sponsored or supported numerous lawsuits that it claims have "arrested the growth of a [problem] that would have grown worse and worse." The suits are not limited to P2P networks. In the last five years, **the RIAA has also sought damages** from more than 30,000 individuals and organizations in the United States suspected of distributing copyrighted works.

The RIAA claims that **pirated content has cost the worldwide music industry as much as \$4.2 billion**, although the claim has been challenged by some researchers.

One project that questioned the RIAA's claims, "**The effect of file sharing on record sales: an empirical analysis**" (Felix Oberholzer and Koleman Strumpf, 2004), found that:

Downloads [of music files] have an effect on sales which is statistically indistinguishable from zero, despite rather precise estimates. Moreover, these estimates are of moderate economic significance and are inconsistent with claims that file sharing is the primary reason for the recent decline in music sales.

The analysis by Oberholzer and Strumpf matched a significant sample of the world's downloads to U.S. sales data for a large number of albums to determine that the sales impact appeared to be minimal.

Since this study, new business models, including the growth of paid-content distribution of individual song downloads, changed the landscape for both music companies and the bands they support. Interestingly, some artists have employed alternative models for distribution and monetization. Coldplay, Radiohead and Nine Inch Nails have experimented with giveaways, asking fans to choose how much they want to pay, and offering premium versions (higher quality audio files) for songs previously released as lower-quality free downloads.

The Motion Picture Association of America (MPAA) has followed a path similar to that taken by the RIAA, pursuing legal actions against a number of P2P sites including 2006 raids at Razorback2 and The Pirate Bay.

Although the actions against Razorback2 resulted in the closure of the site, Pirate Bay came back online only a few days after its raid. At the time of the raid on Razorback2, MPAA Chairman and CEO **Dan Glickman** noted, "Approximately 75 Torrent and eDonkey sites have been shut down in the last year as a result of these efforts." Despite this aggressive approach, P2P sites persist.

After the launch of YouTube and other streaming-video services, television networks initially pressured services whose members had uploaded copyrighted material. Generally, the providers complied, although they resisted calls to proactively screen incoming material to determine ownership or protection as intellectual property. In the last two years, however, some broadcasters have come to see streaming video as an effective tool to promote programs as well as a vehicle to grow traffic on their Web sites.

As a result, actions taken against unauthorized distribution have become less frequent, and many broadcasters have begun offering their own streaming videos of both popular content and newer shows that need to grow an audience. In concert with this trend, broadcasters have joined the music industry in offering individual episodes for sale on services like iTunes, which supports a user's ability to view a program on a computer, a handheld device or (with certain equipment) on a television screen.

2.2. Book Publishing Points of View on Piracy

The Association of American Publishers (AAP) has studied the presence and impact of online piracy for several years. Its Online Piracy Working Group (OPWG) includes more than two dozen representatives from trade, professional and education publishers. In recent years, the AAP has lobbied to move certain countries onto the U.S. trade representative's "watch list" or "priority watch list," heightening the attention paid to international support for intellectual property rights.

The AAP has also supported members who have sought to defend intellectual property rights through lawsuits and the threat of lawsuit. This is most evident in the market for textbooks. In 2008, the AAP applauded a settlement reached by four of its members (McGraw-Hill Education, Pearson Education, John Wiley & Sons and Cengage Learning) against Valore-Books.com, an online textbook marketplace that had been accused of permitting unauthorized third-party sale of pirated educational materials.

Neither the AAP nor its members have pursued the more aggressive paths taken by the RIAA in suing individuals who buy or download unauthorized content. The AAP has collected some data on the extent to which books are available in unauthorized digital form. A May 2008 study, for example, located about 1,100 trade and education titles whose distribution was illegal. This study took place at a time when more than 2 million titles were thought to be in print.

Revenue lost to pirated content is estimated by the AAP at **\$600 million (2006)** and **\$500 million (2007)**. However, these estimates are largely based on the value of printed products created in watch-list countries. The AAP cites "commercial scale photocopying, illegal print runs, unauthorized translations and CD-R burning of text" in explaining the sources of its estimates. While the AAP also cites "massive Internet piracy" as an issue, it is localized to "book and journal publishers doing business" in markets with weak or nonexistent intellectual property protections.

The AAP and individual publishers have focused, appropriately, on the revenues lost when a book is printed illegally in a country with minimal protections for intellectual property rights. Clearly, a pirated book that is printed and sold in place of a legitimate copy costs publishers a sale. Less clear, however, is the impact of pirated content that is not printed and sold.

The presence of 1,100 titles on a mixture of P2P file-sharing sites is, on its own, neither good nor bad news. With more than 2,000,000 titles in print (and a multiple of that number out-of-print), a total of 1,100 titles makes for a very small universe. It is possible that there is significant download activity on this small sample, and if there is, publishers' revenue streams may

be at risk (the threat to academic publishers appears real enough that it has been pursued vigorously, as described above).

When looking at the online presence of unauthorized book content, publishers need to answer three inter-related questions:

1. Are individuals downloading this content at a volume that undermines sales and/or pricing for protected content?
2. Are the individuals substituting this digital version for a printed sale? That is, would they have bought a copy if the download had not been available?
3. Are there any sales spurred by the discoverability of online content that we might otherwise not have gained?

2.3. Some Previous Book Publishing Studies (Strengths and Limitations)

The advent of digital publishing, the ease with which full-book PDFs and various ebook formats can be created, and the growing sophistication of ebook readers have led to a growing amount of digital book content. At the same time, publishers' desire to use digital content to spur online and brick-and-mortar traffic has opened the door to a range of tests seeking to answer the three questions about free content that were posed at the end of the prior section.

Within book publishing, there is disagreement about the extent to which digital distribution harms or helps publishing revenues. There is also uncertainty about the types of books that may be most affected by free access to digital content. Debates about these issues have often been informed by anecdotal evidence or broad comparisons (deciles of titles) conducted over various periods of time. Ultimately, this makes it difficult to compare different tests or experiments and draw meaningful conclusions. It also tends to foster debate about methodology as much as it does the results obtained.

In planning the design of a new set of experiments involving free distribution, we looked at prior work done by O'Reilly Media, Random House and a small

set of examples developed by the Book Industry Study Group (BISG). Designing credible experiments whose results could be analyzed and compared depends on the ability to test enough titles to control for a range of variables that included:

- A variety of markets (in trade, for example, tests in adult, young adult and children's titles) should be tested and measured.
- Although titles with a previous sales history that can be used as a baseline are more credible candidates than front-list books that have just been released, both types of titles should be tested.
- Because there is considerable debate about the impact of free distribution on different types of content, particularly long- vs. short-form, a healthy mix of fiction and non-fiction titles should be tested.
- Author platforms probably matter: a well-known author who is visible across other media (David Pogue, Suze Orman) may view and be affected by free distribution differently than a first-time author with few other marketing options.
- Open files (those without even "light" watermarking) are preferred to less open (trialware, locked) files. The goal of the test is to encourage consumption and hopefully purchase. Preventing recipients from opening a file may prevent a lost sale but it won't encourage a new one.
- The extent to which the availability of free content is actively promoted (as compared with viral means or simply making a file accessible on a public site) may play a role.
- Various formats (PDF, ebook, MP3 and others) may have different impacts on the sale of books.

Taken in sum, it is clear that no one set of tests can cover all of these variables adequately. While the representation by both O'Reilly Media and Random House has started to build an adequate database of tests in certain areas, broader (and structured) participation by a variety of other publishers can only

strengthen and extend the conclusions book publishers can reach when it comes to distribution of free content.

2.4. O'Reilly Point of View and Supporting Actions

For much of the last decade, O'Reilly Media has used its own publishing operations as a vehicle to test and refine its point of view on trends affecting not just O'Reilly but also publishers throughout the industry. In 2002, company founder Tim O'Reilly first published *"Piracy is progressive taxation, and other thoughts on the evolution of online distribution."* Summarized here, the lessons continue to guide O'Reilly in its thinking about the impact of free content:

- Obscurity is a far greater threat to authors and creative artists than piracy, and shoplifting is a bigger threat than piracy. The challenge for publishers and authors has migrated from "getting bought" to "getting discovered."
- Piracy is "progressive taxation." The potential sales loss suffered by the most popular authors is more than offset by increased visibility (and presumably sales) afforded less well-known authors when their content is made available digitally.
- Customers want to do the right thing, if they can. Making content easily accessible does not imply that it will be stolen. Some detrimental piracy may occur, but the average customer wants to recognize good work fairly. Conversely, putting roadblocks in the way of customers who can otherwise be trusted risks undermining the basis of your relationship with them.
- File-sharing networks don't threaten publishing, but they may threaten publishers. The tools themselves are changing the nature of how audiences, content and reputation are aggregated, but publishing (in a growing range of formats) remains. Who will succeed as the aggregators and future publishers is an open question.

- "Free" is eventually replaced by higher-quality alternatives, as we are starting to see in music (DRM-free downloads) and broadcast television (paid, commercial-free downloads of upcoming or just-aired episodes). This may differ from what publishers have tried in the past: charging for content first, or giving content away with the hope that it can later be converted to paid sales.
- There is more than one way to succeed as a publisher. While printing books remains a staple, printing books on its own will not be enough. Ad-supported networks, micro-content sales, subscription models and product sales all may work, or each may work some of the time.

Over the last decade, O'Reilly has tested each of these ideas, evolving as the market and O'Reilly's experience with it direct. In cases where it has pursued unauthorized content sharing, O'Reilly typically starts with a simple take-down request. The firm also tries to understand the results of those who pirate its content: is it popular? If it is, are we (O'Reilly) missing an opportunity to give the customer something he or she wants? In that approach, the number of ways to deliver and monetize O'Reilly content have multiplied.

2.5. A New Model: White-Gray-Back Channel, and Its Implications

In reviewing the history and current sense of free content, we came to see a discrepancy between the binary nature of the free debate (good vs. bad) and the nuances revealed through our own research. Rather than debate "free" as beneficial or harmful, publishers may instead view it as part of a continuum (see *Figure 7, "White-Gray-Back channel: Proposing a more nuanced model"*):

- A "white" market, in which content is created, marketed and sold without the risk or threat of pirated (lost) sales.
- A "gray" market, in which content is used to promote a title or author, with the risk of pirated con-

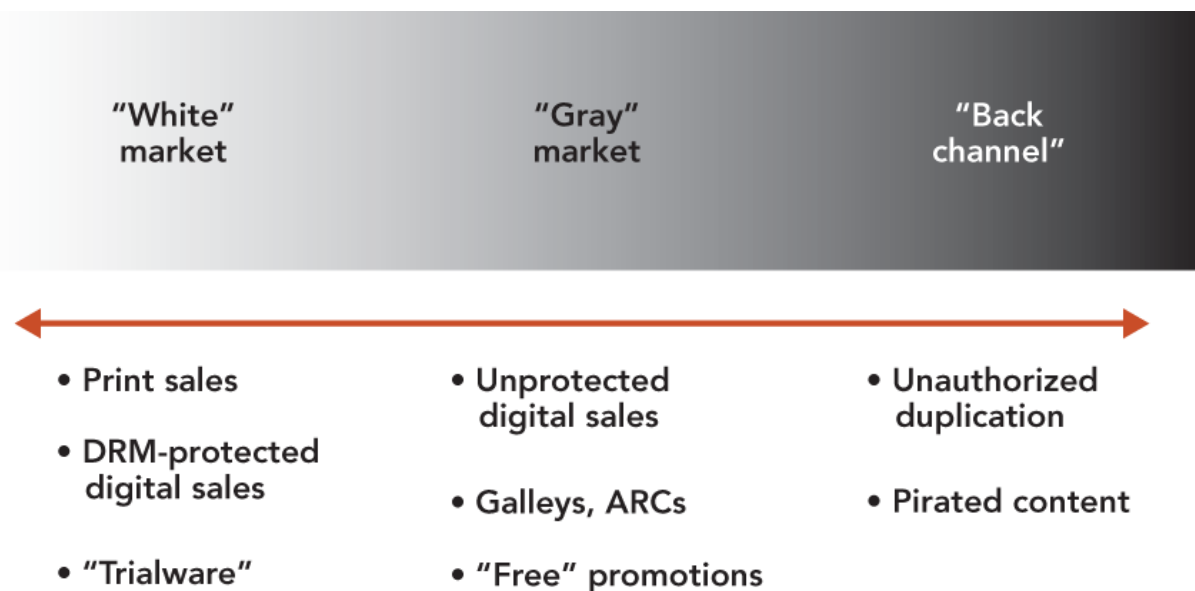


Figure 7. White-Gray-Back channel: Proposing a more nuanced model

tent, but accompanied by a quantifiably better result.

- A "back channel" market, in which content is traded and consumed without fair compensation for its authors or publishers (resulting in lost revenue).

Using these distinctions, publishers have already entered the gray market. Advance reading copies promote new titles but also represent a potential threat to sales. "Search inside" features increase the possibility that a reader may find the critical content and ultimately not buy the book. A free digital download may substitute for a printed copy.

Our initial review suggests that the gray market, on average, can still help grow overall book sales. Increasing the number of people who know about a book, it can be argued, also increases the number who will buy the title, recommend it to others, or purchase other titles in a series.

There are exceptions, such as textbooks, for whom free content is a "back-channel" loss of revenue. It is also possible that gray-channel titles for which free content is a net positive today may become "back-channel" candidates down the road, when digital

readers grow in sophistication or content consumption habits change. With that in mind, a commitment to regularly testing assumptions about free content is in every publisher's better interest.

3. Background on the Peer-to-Peer (P2P) Universe

This section provides an overview of the P2P universe: the recent history, the most significant players, and how the sites currently work, including business models in place at the time of the study.

3.1. Precursors: Early File-Sharing Methods

Usenet and file transfer protocol (FTP), systems that can share files over networks, were established in the 1980s. Usenet and FTP are used primarily for legitimate purposes, although both have also been used for illegal file transfers.

Usenet, a system created by two college students to allow users to read and post articles to others in their newsgroups, operates in a manner similar to today's discussion forums. Rather than store information on a single server, Usenet allows information to be distributed across multiple servers. Users read and post messages on their newsgroup servers, which periodically contact one another to update and exchange information. The introduction of Usenet dramatically reduced server loads and allowed information to be transferred more easily.

Because it is easy to falsify the source information on files transferred through Usenet, and because it is not easy to remove an illegal file from all independent servers at once, the Usenet networks became a conduit for illegal file sharing.

File Transfer Protocol (FTP) is widely used for legitimate file transfers. It can also be employed to transfer copyrighted materials between users. The general layout and construction of FTP makes it unsuitable for

large-scale file sharing, but it became a stepping stone for later P2P networks.

3.2. First-Generation P2P Networks

Characterized by their use of a central server to maintain lists of connected users and available files, the first generation of P2P networks let users search through lists to find files (typically, music) and connected them directly to other users' computers to download the material.

Audiogalaxy started as an **FTP file index in 1999**. It allowed users to connect, download music and chat through an integrated software system. Because Audiogalaxy's copyright infringement filters were insufficient to prevent the exchange of copyrighted materials, the firm was sued by the Recording Industry Association of America (RIAA). Audiogalaxy closed down as a P2P site in 2002 and relaunched as Rhapsody, a streaming music service that works on a monthly subscription system.

Napster, also launched in 1999, is best known for supporting the widespread pirating of music files. Its intuitive interface and constantly updated index system made it easy for users to connect. Napster's service was shut down by **court order in 2000**, pending a copyright infringement trial. In 2001, the 9th Circuit Court of Appeals found Napster guilty of direct infringement, contributory infringement and **vicarious infringement of copyright**. As a result, Napster declared **bankruptcy in 2002**. The Napster name was acquired by Roxio, Inc. and later sold to Best Buy, where it's now associated with a subscription-based streaming music service.

3.3. The Growth of Decentralized P2P Sites

As a result of the Napster litigation, programmers worked to create a P2P client that did not require a central server to transfer or index files. Users could then rely on decentralized networks to share files. The two most successful of these networks were Gnutella and FastTrack, which were accessed by software clients.

P2P network **Gnutella**, which is still active, operates in a manner similar to the original Usenet groups. A user activates client software and connects to a small group of other users (in this case called “nodes”). The nodes connect to one another and distribute addresses for information. The client software caches the addresses in its system (updated whenever a user logs on and periodically while connected).

Whenever a search is performed, the query is sent to those addresses to find a matching file. Each node receiving the query in turn sends the query to all the nodes that it is connected to, creating a “web” of connections that reaches out a certain number of times before the query stops spreading. When a hit is produced, the address information is sent back to the searching computer and a file transfer is attempted between computers.

A client program is required to access the Gnutella network. Limewire and Morpheus are two of the most well-known Gnutella clients.

Limewire is the only P2P software client not to have been shut down by lawsuits. The creators of Limewire learned from the mistakes of other client programs, beginning first by making the program open source and allowing outside developers to play with the code and provide variants and contributions to the program. In addition, Limewire was quick to stop bundling malware and malicious programs with its software (a problem that surfaced early on with client programs). In 2006, as a result of numerous other P2P software firms being taken to court over copyright issues, Limewire implemented filters to curb illegal downloading of copyrighted material. It maintains a copyright policy agreement that all users must agree to before using the software.

Morpheus, another P2P software client, was originally launched in 2000 using FastTrack protocols, but it shut down in 2002 due to licensing disputes. It relaunched in 2003 under Gnutella protocols and continued to function until June 2008, when another lawsuit shut it down again.

Though similar in most respects to Gnutella, the **FastTrack** P2P network does not allow users to search across different network clients (that is, Kazaa users cannot search Grokster files and vice versa). To speed download rates and make the transfer process quicker, FastTrack breaks up files using an approach that makes corruption of files more widespread and harder to identify. In pursuit of download speeds, FastTrack also developed “supernodes” that use nodes with more robust connections to act as proxy servers, regulating data flow and helping make connections between other nodes. As a direct result of FastTrack’s success with download speeds, Gnutella developed and implemented a similar approach.

Once a very popular P2P software client, **Kazaa** is now known for the high amount of spyware and malware that comes bundled with its program. Between 2004 and 2006 Kazaa owners fought legal battles in the United States and Australia over copyright infringement. **The cases were settled out of court.** After providing \$100 million in reparations to music companies, Kazaa stopped updating its site in July 2006 and no longer offers downloads of its program. This has not stopped users from maintaining the program on the FastTrack network, which runs independent of the client programs. As a result, music firms have continued to pursue individual downloaders of music files. Once such case (Oct. 2007) involved Jammie Thomas, a 30-year-old mother who was sued by six major record companies for the illegal distribution of 24 songs. Thomas was ordered to pay \$9,250 for each song distributed as a result of her use of the Kazaa program. In Sept. 2008, judge Michael Davis granted a retrial because of faulty instructions **given to jurors in the first case.**

Similar to Kazaa in function and appearance, P2P software client **Grokster** is notable for legal battles culminating in a Supreme Court decision. In the case

MGM Studios, Inc v. Grokster, Ltd., Grokster tried to use the “Sony safe-harbor principle,” which allows for the sale of copying equipment if the product is widely used for legitimate, unobjectionable purposes. The argument was rejected because the Grokster program was not “widely used for legitimate purposes”—it was estimated by MGM and held in court documents that more than 90 percent of all files downloaded on Grokster were in violation of copyright. As a result of this decision, Grokster is no longer in service.

3.4. How the Current P2P Sites Work (BitTorrent)

P2P networks typically rely on BitTorrent, a file sharing protocol that allows users to simultaneously download parts of a file from multiple users. A user who possesses and uploads a complete file is called a “seeder.” Those acquiring the file are “leechers.” Once a leecher acquires the full file, that user can become a seeder and continue to distribute the file.

A seeder, then, can be either the initial owner or a leecher who has downloaded all of the parts of a seeded file. As long as there is at least one seeder with a full file on hand, the file can continue to be shared. Additionally, if the connected leechers have a full copy of the file between them (called a “distributed copy,” as it is distributed among multiple users), the full file will continue to be seeded.

Even without a full copy available, leechers can continue to obtain pieces of the file as long as someone in the network has them. Because leechers can become seeders, and even an incomplete copy of a file can be a source for downloadable content, simply deleting a seed does not necessarily stop the file from being distributed.

3.5. How Users Work with BitTorrent

Distributing or downloading files from BitTorrent-based P2P networks requires five components: A content file, a torrent file, an index, a tracker, and a client.

Content file

A content file is the file that a user is attempting to acquire (i.e. an ebook, a movie, a software program, etc.). With BitTorrent, a content file is broken

up into segments (“hashes”) that allow the complete content file to be obtained from multiple users simultaneously. This is typically faster and more reliable than connecting to a single user or server to download the entire content file.

Torrent file

A torrent file is required to “seed” a content file through a P2P network. The torrent file acts as a map, outlining the hash configuration and information on how to acquire the file through the P2P network. Torrents themselves are typically only a few kilobytes in size and contain none of the content that a user is trying to acquire. They only guide a client program to the file that is desired.

Index

Though performing two distinct functions, the index and the tracker are increasingly becoming two parts of a collective whole. BitTorrent indexes—also known as BitTorrent search engines—are generally Web sites that index available torrent files for users to download. They have no contact with the content file and only provide a listing of torrent files and where they can be downloaded. Users can search an index for a file and the search results will contain a link to a site that is hosting the torrent file. This is similar to the way Google searches work.

IsoHunt.com is a good example of a strict index-based site. The site catalogs trackers and torrent download sites and provides search capabilities to find torrent files. As of Nov. 2008, isoHunt has indexed more than 400 sites, 70,000 trackers, 1.5 million torrents and 37 million files. Also in 2008, after being sued by the CRIA (Canadian Recording Industry Association), isoHunt filed a countersuit claiming that, as a search engine, the firm has no connection with any copyrighted files and merely serves as a facilitator for the search of freely-distributed torrent files. The results of these legal actions are pending.

Tracker

BitTorrent trackers are communication sites that make the distribution of content files possible.

When you open a torrent file, it will access a tracker that provides a listing of all available seeders and leechers. Again, it should be noted that the tracker does not exchange any content information, it simply provides a listing of peers that a user will access to acquire the content file. A torrent file is needed to access the tracker, and that access is only necessary to initiate a download by providing sources. Once that has occurred, the download can continue without accessing the tracker again. However, it is common practice for a tracker to be accessed periodically to refresh available sources and expedite the download. Trackers generally don't exist independently of indexes, as it is necessary for users to find the torrent file in order to access the tracker.

As the leading BitTorrent tracker and index, *The Pirate Bay* has faced numerous legal threats and actions, including raids on its servers and a *judgment against the site's founders*. The site remains open, even appearing in the news in late 2008 for its *request to have someone donate an Amazon Kindle e-reader to the The Pirate Bay's CEO*. The Pirate Bay is also noted for buying and relaunching Suprnova.org, a search index that had closed in 2004 due to legal issues.

Other trackers of interest are *Mininova*, *BT Junkie*, *Torrent Reactor*, and *FullDLS*. Some trackers have been pursued vigorously in the courts and ultimately forced to shut down. TorrentSpy closed its doors after being ordered to pay \$110 million for copyright infringement. LokiTorrent was sued by the Motion Picture Association of America (MPAA) and subsequently received more than \$40,000 from members to pay for legal expenses.

Client software

BitTorrent client software gives users the ability to seed a file, read a torrent file, and manage file downloads. Some clients also provide download statistics and additional features. Clients include: *Vuze*, *BitTornado*, *Transmission*, *ABC - Yet Another BitTorrent Client*, *µtorrent*, and *Bitcomet*. In addition to common BitTorrent software, various services help content providers track the volume of file

sharing over time. *Baytsp* offers a tracking tool, *EarlyPropagator*, that provides data on shared files.

3.6. A P2P Timeline

Precursors

- 1979-1980—Usenet is created
- 1985—FTP specifications standardized

First Generation

- 1998—Audiogalaxy is launched; Digital Millennium Copyright Act passed
- 1999—Napster launched; First lawsuits against Napster (continue until 2001)
- 2001—*A&M Records, Inc. v. Napster, Inc.* results in Napster being shut down
- 2002—Recording Industry Association of America (RIAA) sues Audiogalaxy

Decentralized Systems

- 2000—Gnutella network created
- 2001—Limewire client released (uses Gnutella network); FastTrack network created; Kazaa client released (uses FastTrack network); Grokster client released (uses FastTrack network); Morpheus client released (originally uses FastTrack network, switches to Gnutella in 2003)
- 2003—RIAA begins suing private citizens for file sharing; RIAA and the Motion Picture Association of America (MPAA) win Supreme Court case against Grokster, site closed down
- 2008—Morpheus client shuts down due to pending lawsuits

BitTorrent

- 2001—BitTorrent created
- 2002—Suprnova.org launches (torrent index)
- 2003—The Pirate Bay launches (torrent index); isoHunt launches (torrent index); TorrentSpy launches (torrent index)
- 2004—LokiTorrent launches (torrent index); MPAA issues cease and desist orders and many small BitTorrent sites close down

2005—LokiTorrent shuts down; Mininova launches (torrent index)

2006—Swedish police raid The Pirate Bay; the service continues to function

2007—The Canadian Recording Industry Association (CRIA), the MPAA and the RIAA continue to issue legal threats toward many BitTorrent sites; police raids continue on servers in a number of countries

2008—TorrentSpy shut down; Italy redirects all Internet traffic within its borders away from The Pirate Bay and forwards it to the International Federation of the Phonographic Industry (IFPI) Web site

3.7. Terms and Definitions

We have tried to consistently use a variety of terms throughout this project. The most critical terms and definitions are outlined in this section. Some terms are not directly referenced within this project, but they are important concepts or topics that deserve mention.

Digital bundles

Multiple formats of digital content for a single title, made available for a single price or sold in concert with a print version of a title.

Digital rights management (DRM)

Technologies or approaches that limit the use of digital content. Light DRM includes approaches like watermarking that help track sources and uses of content. Files that include DRM-free content can be used, re-used, duplicated and/or distributed freely, although there are often restrictions on content use or distribution that the file itself is not configured to enforce.

Distribution

The release of copyrighted content, either in whole or in part, to help drive sales or other revenue opportunities.

Experiments

In the context of this paper, a documented plan to monitor title-specific baseline and subsequent

point of sale (POS) data to establish the impact of freely available digital content on book sales.

Google Book Search

A service provided by Google that allows users to search all of a book's content, browse copyrighted content with the permission of the publisher and download public domain content as a PDF. Provides links to buy or borrow books whose content is available.

Leeches

Users who are actively downloading files that have been made available on a peer-to-peer network.

Peer-to-peer (P2P) networks

Connect ad hoc participants using protocols that support sharing files across the cumulative bandwidth of network participants. The dispersed nature of P2P networks is a key difference from centralized server-based systems.

Piracy

The unauthorized use of copyrighted content.

Platform

In the context of this project, any forum or medium available to an author for discussion or promotion. For example: an author with a blog, weekly column in a newspaper, or a regular appearance on a television program.

Point of sale (POS) data

Information provided by retail accounts that summarize sales for individual book titles. Sales data is generally available weekly; it may also be tracked daily. The data reported by a given account is limited to copies sold within that account.

Protocol

The rules that govern how data exchanges work between two or more machines connected through a network. Usually documented as a standard.

Safari Books Online

A subject-specific subscription service offered by O'Reilly Media and Pearson Education. Safari pro-

vides online access to digital versions of titles from two dozen publishers.

Search Inside the Book

A search feature offered through Amazon that returns search results based on every word inside a book. Similar services are available through other online retailers.

Seeds

Files made available on a peer-to-peer network.

Trialware (also called demoware)

Content generally distributed freely with access restricted by time, output or completeness (a limit on how much content may be read). Trialware content may also use digital rights management to restrict the ability to duplicate files.