2012-3-7

Content Provision Strategies in the Presence of Content Piracy

Monica Johar, *et al.*

Information Systems Research
Vol. 23, No. 3, Part 2 of 2, September 2012, pp. 960–975
ISSN 1047-7047 (print) | ISSN 1526-5536 (online)
http://dx.doi.org/10.1287/isre.1110.0406

Content Provision Strategies in the Presence of Content Piracy

Monica Johar
The Belk College of Business, University of North Carolina at Charlotte, Charlotte, North Carolina 28223, msjohar@uncc.edu

Nanda Kumar, Vijay Mookerjee
Naveen Jindal School of Management, The University of Texas at Dallas, Richardson, Texas 75080
{nkumar@utdallas.edu, vijaym@utdallas.edu}

We consider a publisher that earns advertising revenue while providing content to serve a heterogeneous population of consumers. The consumers derive benefit from consuming content but suffer from delivery delays. A publisher's content provision strategy comprises two decisions: (a) the content quality (affecting consumption benefit) and (b) the content distribution delay (affecting consumption cost). The focus here is on how a publisher should choose the content provision strategy in the presence of a content pirate such as a peer-to-peer (P2P) network. Our study sheds light on how a publisher could leverage a pirate's presence to increase profits, even though the pirate essentially encroaches on the demand for the publisher's content. We find that a publisher should sometimes decrease the delivery speed but increase quality in the presence of a pirate (a *quality focused strategy*). At other times, a *distribution focused strategy* is better; namely, increase delivery speed, but lower quality. In most cases, however, we show that the publisher should improve at least one dimension of content provision (quality or delay) in the presence of a pirate.

Key words: content provision and distribution; delivery delay; content piracy; P2P networks
History: Ram Gopal, Senior Editor; Karthik Kannan, Associate Editor. This paper was received on August 3, 2010, and was with the authors 4 months for 2 revisions. Published online in *Articles in Advance* March 7, 2012.

1. Introduction

The last decade or so has witnessed a tremendous increase in the consumption of (usually free) electronic content over the Internet (Pan et al. 2003). We consider a content publisher (CP) that offers free content to attract traffic and monetizes the traffic in the form of advertising revenue (Gallaugher et al. 2001, Oh 2007). The traffic attracted by the publisher depends on the quality of the content (affecting consumption benefit) and the content distribution delay (affecting consumption cost). Taken together, the publisher's decisions concerning the key traffic influencing attributes of the content (quality and delay) can be referred to as a *content provision strategy*. Choosing an appropriate content provision strategy is a critical issue for most publishers that depend on advertising as the main source of revenue.

The quality of content at a publisher's site is affected by factors such as the staleness of the content, the credibility of the content, other externalities such as the presence of community discussion forums, and so on. In general, quality is any aspect (other than delivery delay) that improves the consumption experience. The other dimension of a content provision strategy affects the delay that consumers experience while consuming the content. High content delivery delays can lead to a permanent loss in traffic for a publisher, and a variety of technologies have emerged to address delay bottlenecks in serving content over the public Internet.

A content delivery network (CDN) is a delivery technology that addresses a publisher's need for the speedy and reliable distribution of content (Vakali and Pallis 2003). CDN providers are firms that centrally manage the delivery of content by replicating or caching content at high-demand locations and dynamically matching user requests for content to locations that can provide the content (Hosangar et al. 2008). Content delivery networks scale well because they are able to achieve long-distance collaboration among servers that replicate the content. On the other hand, peer-to-peer (P2P) networks represent a different, low-cost way to distribute content. The use of P2P networks as a content distribution method extends beyond the exchange of music files over the Internet. Such networks have been used for delivering general-purpose content (including real-time content) on the public Internet (Stolarz 2001, Padmanabhan et al. 2003). Commercial examples include Kontiki (recently acquired by Verisign), which will be used in AOL's In2TV effort to make available for download thousands of old television shows free of charge, and StreamAudio (recently acquired by ChainCast), which uses peer-assisted technology to provide streaming services for radio stations.

This study examines a publisher's content provision decisions in the presence of content piracy. The pirate uses the publisher's content (without appropriate licensing arrangements) to divert consumers away from the publisher's site. This diversion of consumer traffic can have both revenue and cost implications for the publisher. On the other hand, the pirate site may or may not run its own ads to generate revenue, but the costs of content provision are typically much lower because pirates often use P2P technology to distribute the content. In general, although other distribution technologies may be used to distribute pirated content, P2P networks represent an interesting case to consider for several reasons (Julian 2001, Parmeswaran et al. 2001): (1) P2P networks drastically lower the distribution costs for the pirate; (2) they represent a fast and reliable way to distribute content, thus providing consumers with a viable alternative to the official content of the publisher; and (3) P2P-based pirate networks are observable in real-world situations.

To provide an example of the complexity involved in a publisher's content provision decisions, consider the online sports news and entertainment site www.foxsports.com that uses the CDN provider Akamai™ (Eisenmann 2002) to deliver its content. The content is free, so it is reasonable to assume that the main source of revenue for FOXSports.com comes from advertising. However while considering its content provision options, FOXSports.com would need to consider the possibility that some of its consumers could be lured away to access its content via a P2P site (e.g., www.onlinesports24.com). Although the content at Onlinesports24.com (more generally, the consumption experience) is not a perfect substitute, some casual fans could choose to patronize a faster, lower-quality P2P site over a slower but higher-quality "official" site. This loss of traffic to Onlinesports24.com could mean loss of ad revenue for FOXSports.com.[1] Thus the presence of a P2P pirate must be factored by a publisher when developing its content provision strategy. Other examples include P2P networks like Zattoo and SpeedyTV that stream episodes of TV serials and movies. Such content is otherwise freely available on several websites such as Hulu.com, LikeTelevision.com, and SurfTheChannel.com.

A publisher's content provision costs consist of content creation and content distribution costs. Content creation costs consist of a fixed component that depends upon the content quality the publisher chooses to provide. The distribution costs include a variable component (that depends on the number of requests served) and another fixed component that depends on the delay chosen by the publisher. Facing the costs and benefits of content provision, the publisher chooses a delay and quality that maximizes net profit.

We examine how the publisher's choice of optimal delay and quality could change in the presence of a P2P network. With a P2P network, some consumers may choose to obtain the content from a P2P site rather than from a publisher's official site. This has several implications. First, the consumer may suffer some loss of quality at the P2P network. This drop in quality may be due to outdated or incomplete content or from the loss of other benefits (such as a discussion forum run by the publisher). On the other hand, the consumer is likely to incur lower delay when consuming content from the P2P network. Finally, the publisher's advertising revenue is lowered because consumers who consume content from the P2P network do not generate advertising revenue. Because of these effects, the publisher's choice of content quality and delay can be different with and without a P2P network. We show that the publisher should, in most cases, improve at least one of the dimensions of content provision (content quality and delay) with the entry of a P2P network. One of the most interesting findings from this study is how a publisher can leverage the entry of a P2P network to increase profit. Sometimes, the profit increase occurs because of a cost-side effect; i.e., the P2P entry leads to lower distribution costs. For example, when the consumer disutility of delay is low, with the impression-based advertising model cost-per-thousand-impressions (CPM) revenue model, the P2P network *filters out* low valuation consumers and increases the publisher's profit by *reducing distribution costs*. However, in other cases, the P2P entry can also be used to increase the profit by increasing ad revenue, i.e., a demand-side effect. For example, with the action-based advertising model cost-per-action (CPA), the publisher can benefit from P2P entry when advertising price increases because of traffic filtering by the P2P network.

Our paper has some similarities with studies on digital piracy and with classical models of pure competition. However, there are some important differences. In the context of digital content piracy, the positive and negative effects of piracy of digital goods (such as software and music) have been studied in prior research. For example, Bhattacharjee et al. (2007) report that in the context of music piracy, file sharing does not hurt the survival of top-ranked

[1] The OnlineSports24.com site does display ads, but these are its own ads, not those run by FOXSports.com.

albums, but it does have a negative impact on low-ranked albums. On the other hand, they also find that minor labels better utilize file-sharing networks to popularize their albums. Gopal and Gupta (2010) find that in the context of software piracy, bundling can be profitable even when the very act of bundling increases the piracy level of one of the products in the bundle. In these cases, the main source of revenue comes from directly selling the content. However, market demand can expand from positive word-of-mouth effects (e.g., through piracy), better adaptation to technology changes, or from positive externalities of consumption. Such market expansion can help increase the profit from the legitimate sales of a product. In contrast, we are considering a market in which a publisher is distributing general-purpose (free) content and revenue comes from advertisements. Because there are no market-expanding forces in our study, under some conditions the monopolist publisher's profit could increase as a result of piracy. In the context of content competition, a publisher's profits are typically hurt as a result of the entry of a competitor (again, in the absence of market expansion). In contrast, we find that the entry of a P2P pirate (which can loosely be considered competition for the publisher) can sometimes help but at other times hurt the publisher. The main difference is our study is that based on what one can infer from real-world scenarios, we do not consider the P2P pirate to be a strategic entity that optimally chooses content provision parameters (such as quality or distribution speed) in equilibrium to maximize profit. Instead, the P2P pirate is considered an exogenous force acting on the publisher. Thus, the setting in our study (and hence, the main result from the analysis) is different from a classical model of pure content competition.

The rest of the paper is organized as follows. In §2 we present a base model that sets up the analysis in later sections. In §3 we characterize the publisher's equilibrium investment in quality and distribution services with the impression-based advertising revenue model (CPM) and present the main findings. The interplay of market characteristics with the publisher's equilibrium content provision strategies and the impact of the P2P network on the publisher's profits are highlighted. In §4, we revisit the same issues as those in §3 using a CPA advertising model for the publisher's revenue. In §5, we discuss model extensions and provide summary and conclusions in §6.

2. Modeling Consumer Traffic

Consider a market with a publisher whose content is valued by a consumer population. The publisher disseminates the content to consumers using a distribution service, which for the purposes of this study will be assumed to be externally procured from a content distribution provider. In the absence of the P2P network, the content is distributed exclusively via a content delivery network, and consumers interested in viewing the content must visit a site provisioned by the network.

2.1. Quality of Content

With a P2P network, consumers can access content either directly from a CDN site or through a site within the P2P community. However, the consumption experience at these two sources is not identical in terms of the content quality available. Prior research has proposed numerous frameworks for the notion of content quality on the World Wide Web. These include (but are not limited to) currency (up-to-date content), accessibility (availability of content), security (extent to which access to information is appropriately restricted), and credibility (extent to which information is regarded as true or reliable) (Pipino et al. 2004, Yang et al. 2005, Knight and Burn 2005). We expect that the publisher can control some of these commonly accepted dimensions of quality. Therefore, in our model we allow the publisher to optimally choose between offering (i) a low quality option (q_l) or (ii) a high quality option (q_h). The low quality option can be considered as a base level of quality that must be offered by the publisher to continue to be in business. However, when it is more profitable, the publisher may offer content that is of higher quality.[2]

We assume that in comparison, the content at a P2P site is of lower quality (Parameswaran et al. 2001, Wallach 2002, Damiani et al. 2002). The intuition here is that content at the P2P network is often of a *secondhand* nature, periodically "scraped" off the publisher's site. In addition, although some of the same content may be viewed at a P2P site, this content is usually not a perfect substitute in terms of the browsing experience. Consider an avid fan of a particular sport at a FOXSports.com site catering to news about the sport and other related events. The FOXSports.com site could support discussion groups that allow fans to discuss a sporting event while it is in progress, thus providing a superior consumption experience. Such features would usually not be available from the P2P network. The content available at the P2P site may also not be current; for example, in the case of a live event, the P2P site may lag behind an official site (i.e., a publisher site) by several minutes. As another example, consider a website that provides information on the current state of the financial

[2] Here, for mathematical tractability, we restrict CP's choice of content quality to two levels. However, we relax this assumption in §5, where we allow the CP to optimally choose from more than two levels.

markets. Such a site may provide premium content (e.g., streaming content, videos of expert opinions) that is not offered by the P2P site. Finally, consuming the content from the P2P network may be less secure (e.g., it may be infected with malware) when compared to the consuming directly from the publisher. We therefore let q_{P2P} $(<q_l)$ represent the quality of the consumption experience at the P2P network.[3]

Consumers in our model are assumed to be heterogeneous in the value they place on the quality of the consumption experience. This heterogeneity in consumer traffic can have significant implications for content provision. Returning to the sports example, in 2004 FOXSports.com was chosen by Microsoft to exclusively provide sports content for the MSN portal. The Microsoft deal promised a huge increase in traffic for FOXSports.com but it also inherited many less intense sports fans from the vast MSN audience. Essentially, after the Microsoft deal, the FOXSports.com website began to receive at least two different streams of Web traffic: visitors who directly came to the FoxSports.com site and those who were directed to FOXSports.com site from the MSN site. In this example, direct visitors were, on average, more intense sports fans and can be expected to place a higher value on quality. The key challenge for the sports site was to convert casual sports fans to intense ones so that more advertisements could be displayed and drive up revenues. Essentially, FOXSports.com needed to rethink its content provision strategy—entice casual visitors to explore the site by enriching the content, yet support consistent and speedy delivery of the content to a growing audience.

To account for such heterogeneity in the consumer population, we consider the consumer valuation for quality (θ) to be distributed following an unspecified, general distribution ($f(\theta)$; $\theta \in (0, \infty)$). This allows us to derive some structural insights into content provision strategies in the presence of a peer-to-peer network. Later, we also consider a specific case in which consumer valuation is distributed *uniformly* in the interval $[\theta_L, \theta_H]$ to gain additional insights and to study the impact of various parameters.

2.2. Distribution of Content

In deciding whether or not to consume, consumers compare the value of consuming content with the disutility of the delay experienced in accessing the content. The delay experienced by consumers at a CDN site is influenced by the distribution strategy chosen by the publisher (Ercetin and Tassiulas 2005, Pallis and Vakali 2006). For example, in the case of a news site, effective distribution may require some combination of caching and replication of the origin site at selected, high-traffic locations. Here, the delay of content delivery may correspond to the density of replicated servers in a geographical area. As this density increases, the delay decreases. The notion of density also serves as a good mental model in the context of streaming services such as video-on-demand, where a higher number of video servers in the same geographical area would typically translate to superior delay characteristics. On the other hand, distribution of relatively static content typically relies more on caching rather than replication. Here, decreasing the distribution delay may correspond to an Internet service provider (ISP) allocating more caching space for the content belonging to the publisher.

Broadly speaking, although the exact mechanism of how a certain end-user delay is achieved might vary depending on the characteristics of the content being distributed, it is sufficient for our purposes to emphasize that a publisher can reduce the delay by paying a higher distribution cost. For the publisher, controlling distribution delay is important because it affects both the volume and type of consumers that visit CDN sites to browse content and hence the advertising revenue.

2.3. Derivation of Consumer Demand

The publisher can control the delay experienced in browsing content from a CDN site by choosing a variable x that can be thought of as the opposite of delay, or the *speed* of content delivery. We assume that the maximum feasible delay for the publisher to operate is d_c and minimum possible delay is some fraction of the maximum delay given by $d_c(1-x)$. In general, the publisher can choose a speed x to obtain a delay given by $d_c(1-x)$, where $0 \leq x < 1$ is the (normalized) speed of delivery. Thus, the maximum delay d_c is the delay at a normalized speed of zero. On the other hand, the delay experienced by consumers accessing content from the P2P network is denoted by d_p. Because the CP cannot directly control the delay in the P2P network, in our base model, we treat it as a constant.[4] Denoting δ_0 to be the marginal disutility per unit delay incurred when consuming content, the utility (or surplus) derived by a consumer with quality valuation θ in consuming content from a CDN site is given by $U_{\text{CP}} = \theta q_{\text{CP}} - \delta_0 d_c(1-x)$, where $q_{\text{CP}} \in \{q_l, q_h\}$ and

[3] If the quality of content at the P2P site is above q_l but lower than q_h, then the publisher would never be able to choose q_l as its level of content. If the P2P content quality is greater than or equal to q_h, there would be no demand for the publisher's content via a CDN site. This would lead to zero ad revenue and the market for content (including that made available by the P2P network) would collapse. Thus, we only focus on the case where the P2P content quality is below q_l.

[4] In §5, we relax this assumption and consider the case where the P2P network delay improves with the number of consumers that adopt the P2P network.

q_{CP} is the quality of content offered by the publisher. Similarly, the utility derived by a consumer when consuming content from the P2P network is given by $U_{P2P} = \theta q_{P2P} - \delta_0 d_p$.

In the absence of the P2P network, consumers will access content from a CDN site only if $U_{CP} = \theta q_{CP} - \delta_0 d_c(1-x) \geq 0$. This condition states that a consumer of type θ will consume content only if the consumption utility exceeds the utility from outside options (normalized to zero). Thus all consumers with $\theta \geq \theta^*_{CP}$, where $\theta^*_{CP} = \delta_0 d_c(1-x)/q_{CP}$, will consume content from a CDN site. Consumers with $\theta < \theta^*_{CP}$ will not consume content. Thus, in the absence of the P2P network the demand for content is

$$D_{CP} = n \int_{\theta^*_{CP}}^{\infty} f(\theta)\, d\theta = n(1 - F(\theta^*_{CP})), \tag{1}$$

where n denotes the size of the market[5] and $f(F)$ is the probability density function (c.d.f.) of the customer type θ.

Next consider the case in which the content is also available from the P2P network. Consumers will now consume content at a CDN site only if *both* conditions below are met:

$$\theta q_{CP} - \delta_0 d_c(1-x) \geq 0, \tag{2}$$

$$\theta q_{CP} - \delta_0 d_c(1-x) \geq \theta q_{P2P} - \delta_0 d_p. \tag{3}$$

The first inequality ensures (as before) that consumers will visit a CDN site only if the value of the consumption experience exceeds their utility from outside options. The second inequality requires that consumers that visit a CDN site do so because it is in their best interest; that is, the consumption experience at a CDN site offers higher value (or utility) than that at the P2P network. The first inequality requires that $\theta \geq \theta_0(x) = \delta_0 d_c(1-x)/q_{CP}$ and the second inequality requires that $\theta \geq \theta_1(x) = \{\delta_0 d_c(1-x) - \delta_0 d_p\}/\Delta q$, where $\Delta q = q_{CP} - q_{P2P}$ and $q_{CP} \in \{q_h, q_l\}$. The two inequalities together imply that only consumers of type $\theta \geq \theta^*_{P2P} = \max\{\theta_0(x), \theta_1(x)\}$ will visit a CDN site. The demand for content from CDN sites in the presence of the P2P network is therefore

$$D_{CP}(\theta^*_{P2P}) = n \int_{\theta^*_{P2P}}^{\infty} f(\theta) d\theta = n(1 - F(\theta^*_{P2P})). \tag{4}$$

2.4. Content Provision Costs

We next discuss the costs incurred by the publisher in provisioning content to its consumers. The costs of content distribution involve both a fixed component and a variable component. The variable portion of the publisher's cost is a function of the demand reaching the publisher and is assumed to be $C * D(\theta^*)$, where C is the marginal cost of the demand. This form of the cost function reflects pricing models in the content delivery industry and corresponds to the case where the publisher may be procuring content distribution services from an outside provider. We assume that the fixed component of the publisher's costs consists of two terms, (i) $k_0(x^2/2)$, depending on the speed (x) chosen by the publisher, and (ii) $k_1(q^2_{CP}/2)$, depending upon the quality (q_{CP}) chosen by the publisher. The constant k_0 can be interpreted as factor that scales for the size of the market (the larger the market, the larger the cost of increasing the speed). The convex nature of distribution costs reflects existing pricing by commercial CDN providers.[6] Similarly, the constant k_1 is a factor that scales with the quality of the content.

2.5. Advertising Revenue Models

The publisher in our model can influence consumer behavior by choosing the delivery speed as well as the quality of content offered at a CDN site. By increasing the delivery speed or by increasing the quality of content being offered, the publisher can increase consumption utility at a CDN site. This attracts more consumers to consume content via a CDN site. *The natural question arises: why (and when) would the publisher be interested in attracting more consumers to its CDN sites*? In our model, the effect of traffic on advertising revenue, which continues to be the major source of income for most publishers (O'Donnell 2002) is the key consideration for the publisher's content provision strategy. We analyze the publisher's spending on distribution services (to control delivery delay) and content creation (to control quality) under the two disparate pricing models for online advertising (Jacob et al. 2010): the CPM model and the CPA model. In the CPM model, an advertiser pays the publisher based on the number of impressions of an advertisement. In contrast, in the CPA model the payment for an advertisement is based solely on qualified actions such as clicks, sales, or registrations. With these assumptions, the profit of the publisher is the difference between the total revenue and the provisioning costs (both variable and fixed). Given the demand with and without the P2P network, the advertising revenue generated from this demand depends upon the pricing model: CPM or CPA. We will provide details of these pricing models in the next two sections. In §3, we characterize the publisher's equilibrium spending in provisioning content within the CPM regime with and without the P2P network. This allows us to study how publishers should change the content provision strategy in response to the entry of a P2P network. Sometimes the publisher's profits

[5] The units of n can be thought of as the number of consumers per unit time that desire the content if it were made available to them with zero delay.

[6] We also consider linear and concave cost forms in §5.

decrease in the presence of a P2P network, and in these cases the key is to choose a content provision strategy that minimizes this negative impact. At other times, the publisher can exploit the entry of a P2P network to increase profit. In the §4, we consider a publisher's content provision strategy in the presence of a P2P network under the CPA pricing model.

3. Content Provision Under CPM Pricing

We first discuss how publisher revenue is calculated under the CPM pricing model. Next we formulate the publisher's profit function consisting of ad revenue minus the content provision cost. The profit function is analyzed under various situations to provide insights into how the entry of a P2P network might affect the content provision decisions and the profit of the publisher.

3.1. Revenue Under CPM Pricing

According to recent reports, the CPM model still maintains almost a third of the total revenue in the online advertising market (IAB 2011). To account for CPM revenue varying across websites, we assume that the revenue per impression is *linear and increasing* in the average valuation of consumers visiting the site. The intuition is that the price charged by the publisher is a function of the type of consumers it attracts. Given that consumers of type $[\theta_j, \infty]$; $j \in \{\text{CP, P2P}\}$ visit a publisher site, the advertising revenue generated per consumer is assumed to be $\int_{\theta_j^*}^{\infty} \theta f(\theta \mid \theta \geq \theta_j^*) d\theta = (1/(1-F(\theta_j^*))) \int_{\theta_j^*}^{\infty} \theta f(\theta) d\theta$; $j \in \{\text{CP, P2P}\}$. The product of advertising price generated per impression and the traffic yields the total advertising revenue for the publisher with the CPM model.

A case study with a real online advertising company was used to motivate the above pricing scheme.[7] As an illustration of the above pricing scheme, consider an online advertisement for a sports drink product (e.g., Gatorade) at www.foxsports.com vis-à-vis the same advertisement in www.cnn.com. The former has a clientele mix whose average interest in sports related content is higher than that of the CNN.com audience, whose clientele mix is much more diverse. This allows FOXSports.com to command a higher CPM for a sports drink ad relative to CNN.com. A more targeted site often commanding a higher CPM is not limited to online advertising. For example, advertising a financial product is costlier in a more financially targeted newspaper such as the *Wall Street Journal* versus advertising the same product in an outlet that caters to more diverse interests such as *People* magazine.

[7] Private communication with Venkat Kolluri, CEO, Chitika, Inc., 2010. (www.chitika.com)

3.2. Publisher Profit Under CPM

The publisher's objective is to maximize profit by choosing both an optimal speed and the content quality option (low or high). The publisher's profit $\Pi_{\text{CP}}^{\text{CPM}}(x, q_{\text{CP}})$ as a function of the speed and the quality of content chosen by the publisher is given by

$$\begin{aligned}\max_{x,\, q_{\text{CP}}} \; & \Pi_{\text{CP}}^{\text{CPM}}(x, q_{\text{CP}}) \\ &= \text{Ad Revenue} - \text{Variable Distribution Cost} \\ &\quad - \text{Fixed Distribution Cost} - \text{Fixed Quality Cost} \\ &= n \int_{\theta_j^*}^{\infty} \theta f(\theta) d\theta - Cn(1 - F[\theta_j^*]) - k_0 \frac{x^2}{2} - k_1 \frac{q_{\text{CP}}^2}{2} \qquad (5)\end{aligned}$$

where $j \in \{\text{CP, P2P}\}$.

The profit function of the publisher consists of four terms. The first term is the revenue earned by the publisher, which is simply a product of the publisher's demand and price per unit demand (advertising revenue per impression). We have assumed that all visitors to a CDN site are shown ads; i.e., an impression is generated for every visit. If, however, ads are only shown for some visitors but not for others, we only need to suitably scale the value of n, the total demand rate for the publisher. The second term refers to the variable portion of the publisher's cost. The remaining two terms refer to the fixed costs associated with the speed (x) and content quality offered (q_{CP}), respectively. As noted earlier, advertising revenue continues to be a significant source of revenue for most publishers. However, if consumers move away to a P2P site, the advertising revenue to the publisher could reduce.[8]

The key trade-off in the CPM model comes from the positive and negative effects of increasing the speed or the quality of content offered. An increase in the speed by the publisher reduces delay in accessing content from the publisher's site. Similarly, increasing the quality of content offered by the publisher improves the utility derived when consuming content at the publisher's site. Thus, increasing the speed or quality results in increased traffic because consumers with lower valuation for quality find it attractive to visit the site. The increased traffic increases the advertising revenue of the publisher. However, this also increases the variable portion of the publisher's costs because of an increase in demand, as well as the fixed portion of the publisher's costs, which is increasing and convex in the distribution speed and the quality of the content.

[8] The basic idea here can be considered analogous to the P2P site pirating the publisher's content. As will be seen later, depending on the type of advertising revenue model (CPM or CPA), this piracy sometime helps, whereas at other times it reduces, the publisher's profits.

The publisher's choice of speed and quality of content depends on whether or not a P2P network exists in the market. We first study the impact of entry of a P2P network on publisher's profit. Next, we consider how the publisher should optimally change the content provision strategy to accommodate the entry of a P2P network. At the end of this section, to provide deeper insights, we analyze a publisher's content provision strategy for a specific case in which the consumer valuation is uniformly distributed.

3.3. Impact of P2P Entry

We consider the impact of P2P entry on the profit of the publisher and the publisher's optimal adjustment to the content provision strategy (speed and quality of content in the presence of the P2P network). As seen below, the entry of a P2P network typically has a negative impact on the publisher's profit. However, there also exist conditions where the entry of the P2P network increases the publisher's profit by reducing content provision costs. Typically, however, the P2P entry leads to a more aggressive content provision strategy (speed and/or quality increases). Interestingly, it is possible for profits to decrease despite the publisher employing a more aggressive content provision strategy.

3.3.1. Profit Impact of P2P Entry.

Proposition 1. *In the CPM revenue model, the entry of a peer-to-peer network cannot increase the publisher's profit unless the current content provision strategy is minimal (i.e.,* $x^*_{\mathrm{CP}} = 0$ *and* $q^*_{\mathrm{CP}} = q_l$*).*

Proof. See the online supplement.[9] □

The above result is because the demand reaching the publisher always decreases (never increases) with the introduction of the P2P network. In the CPM model, the publisher's revenue is strictly increasing in demand. Hence, the publisher's revenue typically decreases with the P2P entry. In the presence of the P2P network, the publisher is faced with two alternatives: (1) increase either speed or quality to recapture the lost demand, thereby increasing content provision costs, or (2) tolerate the lost demand and accept the lowered revenue. Together, these reactions drive the result that the publisher's profit typically decreases with the entry of a P2P network.

An exception to the above result is the somewhat extreme case where the publisher employs a minimal provision strategy ($x_* = 0$, $q_* = q_l$) before the P2P entry. In this extreme case (and only in this case), the P2P entry could increase the publisher's profit by lowering content distribution costs. A minimal provision strategy may be optimal when the price per impression is so low that increasing revenue by attracting additional demand is not worthwhile because of the associated increase in provision costs. In such situations, the publisher welcomes the entry of a P2P network because it filters unwanted demand and reduces content provision costs. In Corollary 1, we specifically characterize the impact of introducing P2P networks for distributing content on the publisher's market for content and the resulting profit.

Corollary 1. *Depending on the marginal disutility of delay* (δ_0)*, the impact of a P2P network on the profit of the publisher can be summarized using four cutoff values,* $\delta_1 < \delta_2 < \delta_3 < \delta_4$.

1. $\delta_0 > \delta_4$*: there is no market for the content, with or without the P2P network.*
2. $\delta_0 \in (\delta_3, \delta_4]$*: the publisher can make positive profits only in the absence of the P2P network.*
3. $\delta_0 \in (\delta_2, \delta_3]$*: the publisher can make positive profits with or without the P2P network.*
4. $\delta_0 \in [\delta_1, \delta_2]$*: the P2P network helps sustain the market for the content.*
5. $\delta_0 < \delta_1$*: there is no market for the content, with or without the P2P network.*

Proof. See the online supplement. □

When the disutility of delay is high ($\delta_0 > \delta_4$), either the delay must be low or content quality must be high for consumers to derive positive utility while consuming the content. For this to happen, the publisher must invest heavily in content provision. In the range $\delta_0 > \delta_4$, the costs associated with content provision are so high that positive profits cannot be made with or without a P2P network.

When the disutility of delay is lower ($\delta_0 \in (\delta_3, \delta_4]$), the entry of a P2P network threatens the publisher's advertising revenue to the extent that it is impossible to operate in the presence of the P2P network. Here the P2P network acts as a content pirate and its entry destroys the market for content. The intuition for this effect is as follows. If a P2P network is introduced, the demand reaching the publisher decreases. In response, the publisher needs to increase its delivery speed or content quality or both so as to recapture some of the lost demand. Because of this reduced demand and increased costs, the profit of the publisher decreases as the disutility of delay increases. For the range of disutility of delay $\delta_0 \in (\delta_3, \delta_4]$, there is no combination of the speed and content quality for which the publisher can make a positive profit in the presence of the P2P network. However, unlike the range $\delta_0 < \delta_1$, if the P2P network did not exist, the publisher can operate and be financially viable. The range $\delta_0 \in (\delta_3, \delta_4]$ may be viewed as one where the role of the P2P network as a *content pirate* becomes undesirable and excessive.

[9] An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/isre.1110.0406.

As the marginal disutility of delay further decreases ($\delta_0 \in (\delta_2, \delta_3]$), we get a region in which the publisher's network and P2P network coexist. That is, the publisher is able to make positive profits and maintain a market for content provision both with and without the P2P network. In this region the impact of the P2P network can be to increase or decrease the publisher's profits (we will elaborate on this later).

Another interesting region is identified by the marginal disutility of delay range $\delta_0 \in [\delta_1, \delta_2]$. Here the low marginal disutility of delay creates a curious problem for the publisher—one of attracting too much demand—especially from consumers that have a low valuation. Such demand from low valuation consumers is undesirable because it lowers the price that the publisher can charge its advertisers and significantly increases content distribution costs. To filter such undesirable demand, the publisher needs to lower the speed (and thus increase delay) or reduce content quality to reduce consumer utility from deriving content. However, some minimum speed ($x \geq 0$) and content quality ($q \geq q_l$) must be used to provide the content. In the absence of a P2P network, even at the minimum levels of distribution ($x = 0$) and content quality ($q = q_l$), the publisher cannot make positive profits because of the excessive demand from low valuation consumers. With the P2P network, low valuation consumers are *filtered out*; such consumers prefer the lower quality but faster P2P network over the slower but higher quality publisher network. This result is consistent with Proposition 1. If the current content provision strategy is not minimal, the P2P network always decreases publisher's profit. However, at the boundary ($x^*_{\mathrm{CP}} = 0$, $q^*_{\mathrm{CP}} = q_l$), the traffic-filtering feature of the P2P network becomes valuable, and the publisher's profit increases from the P2P entry because it reduces distribution costs.

Finally, in the range $\delta_0 < \delta_1$, the disutility of delay is too low and there is no market despite the helpful role of a P2P network. One could imagine that this is a case for the publisher to explicitly price its content to lower demand.

3.3.2. Impact on Content Provision Strategy.

Proposition 2. *In the CPM revenue model, a publisher will react to the entry of a P2P network by decreasing both the dimensions (speed and quality) of content provision if*

(i) $k_1 < k_1^{\mathrm{crit}}$,

(ii) $f(\theta_1)(\delta_0 d_c/(q_l - q_p)) < f(\theta_0)(\delta_0 d_c/q_h)$, *and*

(iii) $n(\theta_1 - C)f(\theta_1)(\delta_0(d_c(1-x^*_{\mathrm{CP}}) - d_p) \,/\, (q_l - q_p)^2) < k_1 q_l$, *and* $-(\delta_0(d_c(1-x^*_{\mathrm{CP}}) - d_p)/(q_l - q_p))(f(\theta_1) + f'(\theta_1)\cdot(\theta_1 - C)) - (\theta_1 - C)f(\theta_1) \geq 0$,

where $\theta_0 = \theta^*_{\mathrm{CP}}(x^*_{\mathrm{CP}}, q_h^)$, $\theta_1 = \theta^*_{\mathrm{P2P}}(x^*_{\mathrm{CP}}, q_l)$.

Otherwise, a publisher must always increase at least one of the dimensions of content provision (speed or quality) in the presence of the P2P network.

Proof. See the online supplement. □

The publisher can use different content provision strategies to counter the demand encroachment of the P2P network. Typically, the publisher reacts to the entry of a P2P network by increasing at least one of the dimensions of content provision (speed or quality) to recapture some of the lost demand. However, Proposition 2 identifies a condition where the publisher reduces both the speed and the quality in response to the entry of the P2P network and tolerates the lost demand. Note that entry of a P2P network typically increases the valuation of the marginal consumer reaching the CP, i.e., ($\theta^*_{\mathrm{P2P}} \geq \theta^*_{\mathrm{CP}}$). Therefore, an important implication from Proposition 2 is that if the CP chooses to reduce both dimensions in the presence of the P2P, it must be true that in this region ($\theta^*_{\mathrm{P2P}} \geq \theta \geq \theta^*_{\mathrm{CP}}$) the density function ($f(\theta)$) is *decreasing in consumer valuation* (θ). In other words, in this region the distribution function is concave in consumer valuation. When this is true, there are fewer high valuation consumers in the market and the *expected* loss in demand (and revenue) from the P2P's encroachment is smaller. Therefore, the publisher might find it optimal to control costs (by reducing speed *and* quality), especially when fixed costs of distribution are high, and incur additional loss in demand.[10] In all other cases where this specific condition is not satisfied, a publisher should react to the entry of a P2P network by increasing at least one of the dimensions of content provision (speed or quality).

3.4. Uniformly Distributed Consumer Valuation

To derive some additional insights, we conduct a more detailed analysis of the CPM revenue model when the consumer valuation is distributed *uniformly* in the interval $[\theta_H, \theta_L]$. When the valuation of consumers is uniformly distributed, we find that a publisher should react to the entry of a P2P network by increasing at least one of the dimensions of content provision (speed or quality). The exact content provision strategy following the P2P entry depends on three thresholds ($k_{1\mathrm{crit}}^{\mathrm{CP}}, k_{1\mathrm{crit}}^{\mathrm{P2P}}, k_{0\mathrm{crit}}$) that can be calculated using the parameters of the problem. The first threshold ($k_{1\mathrm{crit}}^{\mathrm{CP}}$) represents the highest value of the marginal cost of quality (k_1) at which the publisher will choose to provide high quality content in the absence of the P2P network. The second threshold ($k_{1\mathrm{crit}}^{\mathrm{P2P}}$) is conceptually similar and governs the publisher's choice of high quality in the presence of a P2P network. Finally, the third threshold ($k_{0\mathrm{crit}}$) is a

[10] For illustrative purposes, we find such a region for the case where consumer valuation follows a triangular density function.

$$f(\theta) = 4(\theta - \theta_l)/(\theta_h - \theta_l)^2 \quad \forall\, \theta_l \leq \theta \leq (\theta_h + \theta_l)/2,$$

$$f(\theta) = -4(\theta - \theta_h)/(\theta_h - \theta_l)^2 \quad \forall\, (\theta_h + \theta_l)/2 \leq \theta \leq \theta_h.$$

threshold for the fixed cost of content distribution (k_0) and governs the publisher's decision to increase or decrease delivery speed.

Corollary 2. *In the CPM revenue model, when the consumer's valuation follows a uniform distribution* ($\theta \in U[\theta_H, \theta_L]$), *a publisher should accommodate the entry of a P2P network by modifying its optimal content provision strategy as follows.*

(a) *If the marginal cost of quality satisfies* $k_{1\text{crit}}^{\text{CP}} \le k_1 \le k_{1\text{crit}}^{\text{P2P}}$ *and the marginal cost of increasing the delivery speed satisfies* $k_0 \ge k_{0\text{crit}}$, *the publisher should increase the quality of content but decrease the speed. This is a* quality-focused strategy.

(b) *If the marginal cost of quality satisfies* $k_{1\text{crit}}^{\text{P2P}} \le k_1 \le k_{1\text{crit}}^{\text{CP}}$, *the publisher should increase the speed and lower the quality. If the marginal cost of quality satisfies* $k_1 < k_{1\text{crit}}^{j}$ *or* $k_1 > k_{1\text{crit}}^{j}$, $j \in \{\text{CP}, \text{P2P}\}$, *the publisher should increase speed at the same quality. These are* distribution-focused strategies.

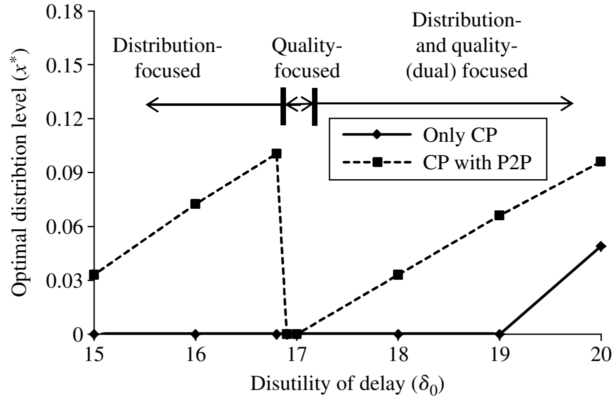
(c) *The publisher should increase both the quality of content and the speed if* $k_{1\text{crit}}^{\text{CP}} \le k_1 \le k_{1\text{crit}}^{\text{P2P}}$ *and* $k_{0\text{crit}} > k_0$. *This is a* dual-focused strategy.

Proof. See the online supplement. □

We delineate the publisher's content offering strategy in the presence of the P2P network into three distinct regions. For example, when the fixed cost of distribution is high ($k_0 \ge k_{0\text{crit}}$), it is more profitable for the publisher to counter the demand encroachment by the P2P network by increasing the quality offered but reducing the speed (*the quality-focused strategy*). On the other hand, when the fixed cost of quality is sufficiently high (k_1) the publisher should increase only the speed without changing the quality (or even reducing the quality) in the presence of the P2P network (*the distribution-focused strategy*). Finally, under situations where the competitive pressure exerted by the P2P network is sufficiently high, the publisher is forced to increase both the speed and the content quality to compensate for the demand lost to the P2P network (*the dual strategy*). An important result from Corollary 2 is that the publisher, for the specific case of uniformly distributed consumer valuation, should never reduce both the speed and the quality in response to the entry of the P2P network. Next, with a help of some numerical examples, we illustrate these different regions associated with the impact of a P2P network on the publisher's market for content.

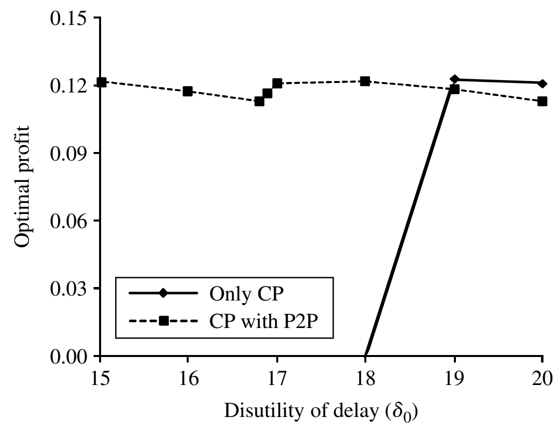
3.4.1. Numerical Illustrations. In Figure 1, we illustrate the boundary effect where the P2P network can increase profit for the publisher as a function of the disutility from unit delay (δ_0). In the entire interval ($\delta_0 < 19$), without the P2P network, the disutility of delay in the publisher network is low enough

Figure 1(a) **Optimal Distribution (x^*) as a Function of Delay Sensitivity δ_0[a]**



[a]Model parameters (Figure 1): $\theta_h = 10$, $\theta_l = 5$, $d_c = 1$, $d_p = 0.3333$, $C = 9.5$, $k_0 = 2$, $k_1 = 0.001$, $q_h = 2.2$, $q_l = 2$, $q_{\text{P2P}} = 1$

Figure 1(b) **Publisher Profit as a Function of Delay Sensitivity δ_0**



that the publisher attracts a large demand even when both the distribution speed ($x_{\text{CP}}^* = 0$) and the content quality offered ($q_{\text{CP}}^* = q_l$) are at a minimum (refer to Figure 1(a)).

In this region, the introduction of the P2P network filters some of the demand, particularly the consumers that have lower marginal valuation of content, from reaching the publisher. Note that in the interval $\delta_0 < 18$ (in Figure 1(b)), a publisher's profits in the absence of the P2P network are zero, implying that there is no market for content. However, in the presence of the P2P network the profits are high enough to offset the costs, suggesting that under these conditions the presence of the P2P network is necessary for a publisher to sustain a market for content, ($\delta_0 \in (\delta_1, \delta_2]$ in Corollary 1). Here the P2P network plays the role of a *creative content pirate* and is necessary to sustain the market for content.

Note that in Figure 1(a), we are also able to illustrate the impact of the P2P network on the publisher's content provision strategy along the lines of Corollary 2. Initially ($\delta_0 < 16.8$) the publisher follows a distribution-focused strategy in response to entry of the P2P network. Particularly, the publisher chooses to offer the same (low) quality content but increases speed in the presence of the P2P network. Eventually in the region $\delta_0 \in (16.8, 17)$, the disutility of delay is high, so the publisher reacts to the presence of the P2P network by offering high quality content. However, in this region the speed remains unchanged with and without the P2P network (i.e., quality-focused strategy). Finally for $\delta_0 > 17$, disutility of delay is high enough that in response to P2P entry the publisher adopts a dual strategy (increasing both speed and quality).

In Figure 1, we identify another region ($\delta_0 \in [18, 18.8]$) where the publisher's profit, in the absence of the P2P network, is increasing in the marginal disutility of delay. When the disutility of delay decreases, the publisher would prefer to invest less in distribution services, but given that the speed is already at a minimum, this is not feasible (refer to Figure 1(a)). As the marginal disutility of delay increases, the distortion between the optimal speed and the constrained speed (speed must be greater than or equal to zero) decreases, leading to a positive impact on profits. Note that in this region, $x^*_{\text{CP}} = 0$ and $q^*_{\text{CP}} = q_l$ such that the publisher's fixed costs are unaffected. Although the demand is adversely affected with an increase in δ_0, the average valuation of the consumers that visit the publisher network increases and hence the advertising price per impression also increases. Furthermore, with this reduction in demand the variable costs of distribution are reduced and as a result the positive effects (higher price and lower variable cost) dominate the negative effect (reduced demand) and the profits increase with δ_0. However, in the region $\delta_0 > 19$ the publisher's distribution exceeds the minimum level ($x^*_{\text{CP}} > 0$; see Figure 1(a)). In addition to these three effects (on demand, price, and variable costs) discussed above, an increase in the marginal disutility of delay also increases the fixed costs (because speed is increasing in δ_0). The negative effect on demand and fixed costs dominates the positive effect when $\delta_0 > 19$. Hence the profits decrease in δ_0.

In Figure 1, the region $\delta_0 \in [18, 18.8]$ has an interesting interpretation. Note that the publisher's profit in this region is nonnegative with and without the P2P network. Therefore, this region illustrates $\delta_0 \in (\delta_2, \delta_3]$ in Corollary 1, where the publisher has a viable market for content with and without the P2P network. However, the traffic filtering effects of the P2P network help the publisher to increase its profit. The intuition is that under these conditions, the coexistence of the P2P and the publisher helps segment the market. This is analogous to the effect of lowering the speed as δ_0 decreases to deter the low valuation consumers from visiting the publisher's site. Thus in the region $\delta_0 \in [18, 18.8]$, the P2P network continues to act as a creative content pirate, although its presence is not necessary to sustain a market for content.

When the disutility for delay increases beyond a certain level ($\delta_0 > 18.8$), the role of the P2P network becomes one of a destructive content pirate. In this region, the presence of the P2P network reduces the publisher's profits.

3.4.2. Impact of Consumer Heterogeneity on Publisher Profit.

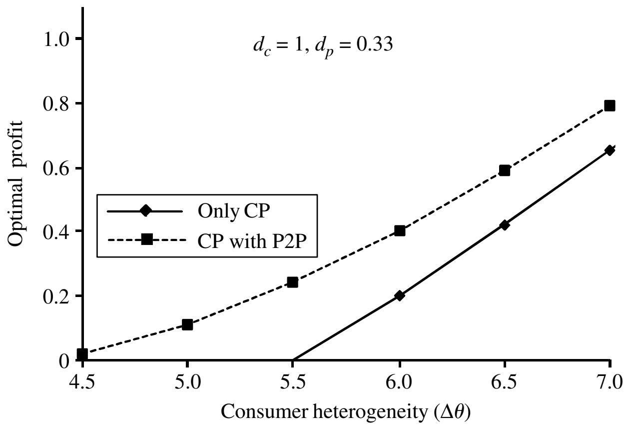
LEMMA 1. *The content provider's profit is increasing in consumer heterogeneity*[11] ($\Delta\theta$) *in the presence or absence of the P2P network.*

Lemma 1 investigates the impact of consumer heterogeneity ($\Delta\theta$) on the publisher's profit. Interestingly, we find that optimal profit of the publisher always increases with consumer heterogeneity. The intuition is that for low values of consumer heterogeneity, most consumers either consume or do not consume. This limits the ability of the publisher to segment the market using speed or quality. On the other hand, at higher levels of consumer heterogeneity, the publisher can use speed and quality to attract the desired part of the market. When consumer heterogeneity is low, the role of the P2P network as a creative content pirate (thus providing an exogenous segmentation device) becomes especially helpful. This can be seen in Figure 2(a), where the profit of the publisher is strictly increasing in $\Delta\theta$ but the beneficial impact of the P2P network is greater when the heterogeneity is low. In Figure 2(a), for small values of consumer heterogeneity (4.5–5.5), we find that the P2P network acts as a filtering mechanism, diverting some of the excess demand away from the publisher. Thus, in this region, the presence of the P2P network helps create a market for content distribution: it is a creative content pirate. Beyond this region, the segmentation benefits of the P2P entry steadily reduce.

Figure 2(b) illustrates that at low consumer heterogeneity, the entry of a P2P network can also be destructive. This phenomenon is revealed when we examine the market from the lens of the *relative* delay in the publisher network (i.e., the ratio the publisher delay over the P2P delay, d_c/d_p). Consider the case where the relative delay increases from (1/0.33) in Figure 2(a) to (2/0.15) in Figure 2(b). In Figure 2(a), where the difference in delay is low, the publisher is better off with the P2P network because the filtering it provides offsets the demand encroachment that

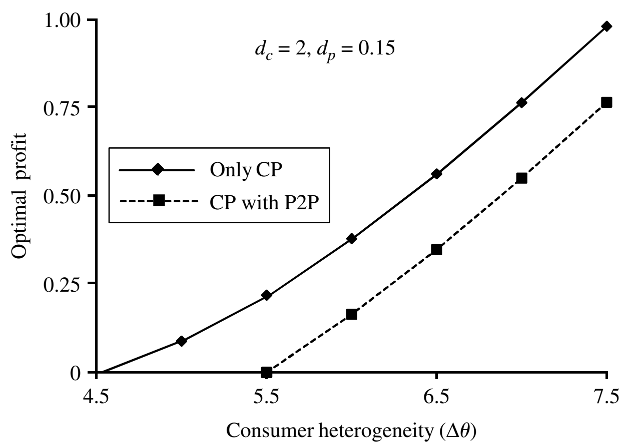
[11] While increasing consumer heterogeneity, we imply increasing the variation in consumer type but not increasing the average consumer type—i.e., increasing $\Delta\theta = (\theta_H - \theta_L)$ while keeping $(\theta_H + \theta_L)/2$ constant.

Figure 2(a) **Publisher's Optimal Profit vs. Consumer Heterogeneity at Low Relative Delay**[a]



[a]Model parameters (Figure 2): $(\theta_h+\theta_l)/2 = 7.5$, $\delta_0 = 16.9$, $C = 9.5$, $k_0 = 2$, $k_1 = 0.001$, $q_h = 2.5$, $q_l = 2$, $q_{P2P} = 1$.

Figure 2(b) **Publisher's Optimal Profit vs. Consumer Heterogeneity at High Relative Delay**



accompanies it. On the other hand, when the difference in delay is large (Figure 2(b)), the P2P network's filtering benefits are not sufficient because it competes away a majority of the publisher's demand. Here, the P2P network destroys the market for content provision and the publisher is better off in the absence of the P2P network.

3.5. Summary

The findings in this section can be summarized as a cost-side effect associated with the entry of a P2P network. Under various situations, the P2P network was shown to have beneficial or harmful effects on the cost of content provision. On the other hand, the entry of a P2P network always led to lower revenues (never higher) for the publisher. Thus when the publisher's profits increased, it was only because the cost of content provision reduced sufficiently to compensate for the lower revenue. In the next section we re-examine these phenomena in the context of a different ad revenue model, specifically the cost-per-action, or CPA, model. The main goal is to study the impact of the P2P entry to see if there is an associated demand-side effect, namely, we study the question: can the publisher's revenue increase following the entry of a P2P network?

4. Content Provision and Profit Under CPA Pricing

We first describe the process of revenue calculation under CPA pricing. Next, we present the publisher's profit function and analyze it to derive insights. The goal is to highlight how content provision and profit change under CPA pricing.

4.1. Advertising Revenue Under CPA

In the CPA model, the publisher only earns advertising revenue based on the number of qualified actions (such as ad clicks, registrations, or sales). The key difference in the CPA model (as compared to CPM) is that publisher's revenue is dictated by the qualifying action rate rather than the demand rate. Although the action rate can be improved by increasing the number of consumers reaching the publisher's site, this strategy has diminishing returns because the low valuation consumers are less likely to generate a qualifying action. The second distinction is that the qualifying action rate is strictly less than the demand because every visit does not translate to a qualifying action. Therefore, the price per action is typically higher than price per impression. That is, there is a price premium ($k_2 \geq 1$) such that the *price per action* $= k_2 *$ *price per impression* is *linear and increasing* in the average valuation of consumers visiting the site. Therefore, as in CPM pricing, the price per action is linear and increasing in the average valuation of consumers visiting the site.

The above pricing scheme has been constructed based on actual pricing models used in the online advertising industry. Experience with online advertising indicates that the "fit" of the content with an ad is an important determinant of the interest shown in the ad by a visitor. In our model, the consumer's interest in the content is modeled by the consumer type construct, θ. In the CPA model, the ad revenue earned by the publisher is predicated on qualified action events. Hence, under the CPA model, it becomes more important for the publisher to select advertisers that offer products or services that match the content being served. Thus, consumers with higher interest or higher valuation for the content will be more likely to generate actions than those that are less interested in the content.

We assume that the likelihood of a consumer generating a qualified action depends on the consumer type (θ). Specifically, the probability density that a consumer of type θ will generate a qualified action (e.g., a click) is assumed to be a power distribution, $A/(B-\theta)^2$, where A and B are scale and shape parameters. The power law distribution is often used to capture phenomena where the quantity of interest can change (increase or decrease) rapidly across items (or segments) in the population, e.g., the popularity of songs, hit-rate of items in a cache, etc.

In our problem, empirical evidence suggests that the click probability can change rapidly across different visitor types in a population. To validate this claim, we collected data on the click events (zero for no-click, one for click) of visitors ($N = 40{,}000$) to a variety of websites (publishers) managed by the online advertiser www.chitika.com. Using logistic regression, we then developed a predictive model that provided a visitor's probability of a click, given profiling attributes obtained from the visitor's cookie data (previous search history, click events on similar sites, etc.) and other data extracted from the http header (e.g., search string, browser, date/time), providing a total of 35 attributes to describe a visitor. Having obtained the click probability values, we used k-means clustering ($k = 4$) to cluster the visitors to the site based on profile attributes. Finally, we obtained the mean click probability in each cluster: [Clus_1 = 0.0001; Clus_2 = 0.002; Clus_3 = 0.01; Clus_4 = 0.26]. Consistent with a power law distribution, the click probability (approximately) increases by an order of magnitude across clusters, thus indicating a geometric progression rather than a linear progression. Using the power law, the expected number of qualified actions can be calculated as below:

$$n\int_{\theta_j^*}^{\infty}\frac{A}{(B-\theta)^2}f(\theta)d\theta, \quad j\in\{\text{CP, P2P}\}. \tag{6}$$

The total expected revenue is, of course, the price per action times the expected number of qualified actions.

4.2. Publisher Profit Under CPA

In the CPA model the publisher's objective is same as in the CPM model, i.e., to maximize profit by choosing the optimal speed and quality of content. The publisher's profit $\Pi_{\text{CP}}^{\text{CPA}}$, as a function of the speed and content quality, is given by

$$\begin{aligned}\max_{x,\,q_{\text{CP}}}\ & \Pi_{\text{CP}}^{\text{CPA}}(x, q_{\text{CP}})\\ &= \text{Revenue per Action} \times \text{Action Rate}\\ &\quad - \text{Variable Cost} - \text{Fixed Costs}\\ &= \left(\frac{k_2}{1-F(\theta_j^*)}\int_{\theta_j^*}^{\infty}\theta f(\theta)d\theta\right)\left(n\int_{\theta_j^*}^{\infty}\frac{A}{(B-\theta)^2}f(\theta)d\theta\right)\\ &\quad - Cn\int_{\theta_j^*}^{\infty}\frac{A}{(B-\theta)^2}f(\theta)d\theta - \left(k_0\frac{x^2}{2}+k_1\frac{q_{\text{CP}}^2}{2}\right)\end{aligned} \tag{7}$$

where $j\in\{\text{CP}, \text{P2P}\}$.

The profit function consists of three terms; the first term is the revenue earned by the publisher, which is simply a product of the action rate generated by the consumers and the revenue earned per action, i.e., the price per action charged by the publisher. In this model, the revenue generated is a function of the number qualified actions (e.g., ad clicks, leads, sales). The remaining two terms represent the variable and fixed costs incurred by the publisher as discussed earlier (identical to the CPM model in §3). For a detailed derivation of the objective function see the online supplement.

The key tradeoff in the CPA model comes from the positive and negative effects of increasing the publisher's speed or content quality. An increase in either speed or quality increases the consumer's utility from consuming content, resulting in a decrease in the valuation of the marginal consumer (θ^*). This has the positive effect of increasing the action rate for the publisher, which is decreasing in θ^*. On the other hand, increasing speed has a negative effect of decreasing the publisher's price per action because an increase in speed increases the lower valuation consumers reaching the CP. In addition, there is the negative effect of higher fixed and variable costs associated with increasing speed or quality.

4.3. Model Analysis

In this section we analyze how the two advertising revenue models compare in terms of the effect of the P2P network on the publisher's profit. Such a comparison is presented in Proposition 3 below.

Proposition 3. *In the CPA revenue model, the publisher's revenue (and profit) may increase because of the entry of a P2P network.*

Proof. See the online supplement. □

Unlike CPM pricing, in the CPA pricing model, the advertising revenue may increase or decrease with demand. Thus, the introduction of the P2P network and the consequent demand encroachment may actually *benefit the publisher's revenue and profit* (i.e., a demand-side effect). In contrast, we have already noted that under the CPM model, the publisher's profit always decreases because of lower revenue (except for the boundary case $x_{\text{CP}}^* = 0$, $q_{\text{CP}}^* = q_l$). The underlying intuition is that the revenue under the two advertising schemes responds differently to the consumer type. Under CPM pricing, the revenue is strictly increasing in demand, so more consumers means more revenue (irrespective of consumer type). In contrast, in the CPA model, a publisher's revenue is dictated by the qualifying action rate rather than demand. Thus, higher valuation consumers, who are more likely to generate actions, are more valuable. Therefore, in the CPA model (unlike CPM), the filtering of low valuation consumers by the P2P network can increase a CP's profit and revenue.

4.4. Uniformly Distributed Consumer Valuation

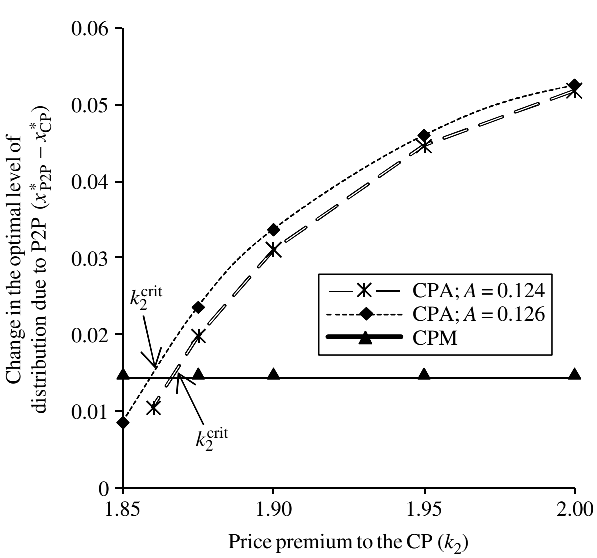
As in §3.4, to gain additional insights, we consider the CPA pricing model for the specific case where consumer's valuation is uniformly distributed ($\theta \in U[\theta_H, \theta_L]$). We find that similar to the CPM model, the publisher's choice of the speed and content quality offered will depend on whether or not the P2P network is present. For brevity, we limit our analysis to how the new (CPA) model parameters, such as price premium and probability of generating a qualifying action, impact the publisher's content provision strategy (and draw contrasts against the CPM model). The key findings of this analysis are summarized below.

PROPOSITION 4. *The change in speed resulting from the entry of the P2P network* ($x^*_{\text{P2P}} - x^*_{\text{CP}}$) *is lower in the CPA model* (*as compared to CPM model*) *if the price per action* (k_2) *is less than a critical value* k_2^{crit}. *Otherwise, the speed difference is higher in the CPA model.*

PROOF. See the online supplement. □

We define k_2^{crit} as the value of k_2 for which the change in speed due to entry of the P2P network ($x^*_{\text{P2P}} - x^*_{\text{CP}}$) is same for the CPM and the CPA model. Proposition 4 outlines the impact of introduction of the P2P network on the publisher's speed as a function of the price premium in the CPA model. In the CPA model we find that the impact of introducing the P2P on the speed depends upon the price premium (k_2). Particularly, Proposition 4 states that the speed impact of introducing the P2P network in the CPA model increases with k_2 (see Figure 3). This is

Figure 3 Impact of P2P Network on the Speed Difference: CPM versus CPA, as a Function of the Price Premium (k_2) of the CPA Model



because at low values of the price premium, the publisher (rather than the advertiser) takes most of the risk of attracting consumers who are likely to generate qualifying actions. As a result for small k_2, the publisher finds it optimal to improve the valuation (type) of consumers rather than simply increasing demand (which also increases the publisher's costs), so the publisher does not try to aggressively recover the lost demand in the presence of the P2P network; i.e., the speed difference $x^*_{\text{P2P}} - x^*_{\text{CP}}$ is small. On the other hand, as the price premium increases, the risk to the publisher from low valuation consumers reduces. As a result, the publisher finds it optimal to increase demand even at the cost of lowering the marginal consumer valuation. Thus, in the presence of the P2P network, the publisher tries harder to recapture lost demand, and the speed increases with the price premium k_2. In Figure 3, we observe that for $k_2 > k_2^{\text{crit}}$, the impact of the P2P network on the speed difference is higher in the CPA model. We next introduce Lemma 2, which concerns the critical value of the price premium, k_2^{crit}.

LEMMA 2. *The critical value of the price premium* (k_2^{crit}) *decreases in the probability of generating a qualifying action.*

PROOF. See the online supplement. □

As the probability of generating a qualifying action increases, the likelihood that a consumer of type θ will generate ad revenue for the publisher increases. This relieves some of the pressure the pressure faced by the publisher to attract only high valuation consumers. Therefore, if the probability of generating a qualifying action is high, the publisher tends to focus more on regaining lost demand in the presence of the P2P network (and is less concerned about lowering the marginal consumer valuation). This effect is illustrated in Figure 3, where the value of k_2^{crit} is lower when the probability of generating a qualifying action (A) increases.

Next we compare how the publisher's choice of offering low or high quality content differs between the CPM and CPA model. Recall that k_1^{crit} represents the critical value of the fixed cost associated with quality; when $k_1 > k_1^{\text{crit}}$, the publisher finds it optimal to offer low quality content. In Figure 4, we plot k_1^{crit} against the price premium (k_2) in the CPA model. Obviously, for the CPM model, k_1^{crit} is independent of k_2 (see Figure 4). Interestingly, we observe that k_1^{crit} in the CPA model is increasing in the price premium. This implies that the lower (higher) the price premium, the more likely the publisher is to offer low (high) quality content. Again, the intuition is closely related to the discussion of Proposition 4. At low values of the price premium in the CPA model, the publisher must attract only high valuation

Figure 4 Impact of P2P Network on the Quality of Content: CPA versus CPM, as a Function of the Price Premium (k_2) of the CPA Model

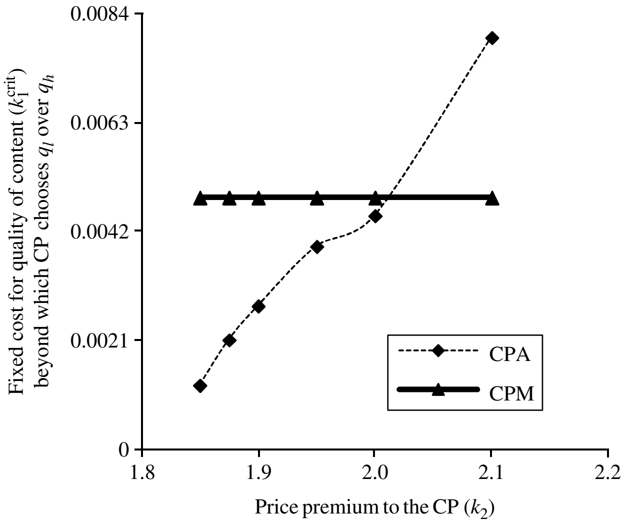
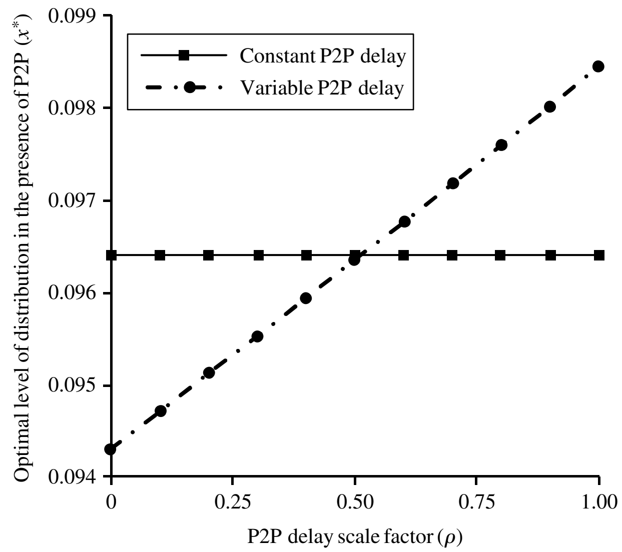


Figure 5 Impact on Publisher's Level of Distribution When P2P Network Delay Reduces with Number of Peers



consumers. This is achieved by offering low content quality and ensuring that only high valuation consumers consume at the publisher's site. Another way of interpreting the findings in Figure 4 is that for low (high) values of price premium, the publisher is likely to choose lower (higher) content quality in the CPA model as compared to the quality chosen in the CPM model.

It should be noted that although we do not explicitly restate the results, we find that in the CPA model (as in the CPM model) the introduction of the P2P network, under certain conditions, can have the effect of destroying the market for content. On the other hand, under certain conditions, it can also have the effect of creating a market for content that was otherwise infeasible when the publisher solely provides distribution services.

5. Model Limitations and Extensions

5.1. Variable P2P Network Delay

Recall that in our basic model we considered the delay in the P2P network to be constant. In this model extension, we allow the delay in the P2P network to reduce with the number of peers consuming content. Particularly, the delay in the P2P network is $d_p^{\max}(1-\rho(F[\theta_{\mathrm{P2P}}^*]-F[\theta_p]))$, where $d_p^{\max} > d_p$. Here, θ_{P2P}^* is the marginal consumer reaching the publisher's site in the presence of the P2P network, and θ_p is the marginal consumer consuming content from the P2P network. Hence, $F[\theta_{\mathrm{P2P}}^*] - F[\theta_p]$ represents the proportion of consumers that consume from the P2P network. The delay in the P2P network decreases with the number of P2P consumers. The delay factor $\rho(\in [0, 1])$ controls the impact of P2P adoption on P2P delay. The variable delay model provides results that are similar to those when the P2P delay is constant. However, the competitive pressure imposed by the P2P network with variable delay can be lower or higher than that with constant delay. This can be seen in Figure 5: when the delay scale factor is relatively low, the distribution chosen by the publisher is also low. The reverse is true when the impact of adoption on delay is relatively high.

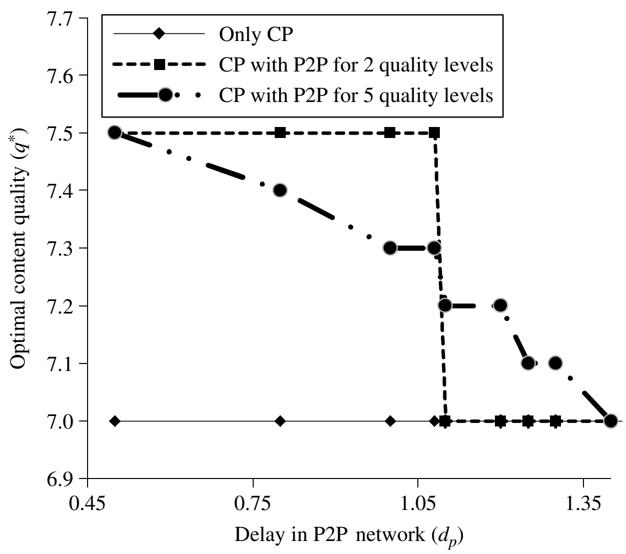
5.2. Multiple Content Quality Options

In our basic model, for mathematical tractability, we allow the publisher to choose from two levels of content quality as a part of the overall content provision strategy. In our model extension, we consider the case where the publisher can choose to offer several (up to five) levels of content quality. Again, we find the qualitative nature of the results to be similar to the case where the publisher could choose from two quality levels. However, as expected, as the number of quality options increases, the publisher finds it optimal to gradually change quality (i.e., there are more switch points). This is illustrated in Figure 6.

5.3. Alternative Forms for Distribution Costs

So far we have considered the publisher's fixed distribution costs to be increasing and convex in distribution speed. Convex costs, of course, guarantee that the profit function is concave in distribution speed. However, there is no inherent restriction in the model limiting the form of the cost function. In the specific model, for *linear costs* all the results hold because profit function continues to be concave in the speed. However, for the case of some *concave* cost function, $\partial^2 k(x)/\partial^2 x \geq -(n\delta_0^2 d_c^2/\Delta\theta q_{\mathrm{CP}}^2)$ is a sufficient condition

Figure 6 Impact of Increasing the Number of Quality Options on Publisher's Content Provision Strategy



to ensure that the profit function is concave. That is, the cost function should not be strongly concave in the speed. Convex or linear costs appear more reasonable in our problem because increasing the distribution speed should at least require a linear increase in the distribution effort (and cost).

6. Summary and Conclusions

We set ourselves the task of examining the impact of P2P distribution networks on a publisher's content provision decisions: (1) the speed and (2) the quality of the content. We presented a model in which a heterogeneous population of consumers views content (created by a publisher) from either an official source (managed by a content distribution provider) or from a source within a P2P network. Generally speaking, those consumers who place less value on the quality of the content (and the associated experience of viewing it at a website) prefer to get the "bare bones" content from the P2P community. Some consumers (who place a higher value on quality) prefer to view the content by accessing an official source that publishes it. The publisher earns revenue from advertising that depends on the number and average valuation of consumers that view content from an official source.

Although P2P networks have typically been viewed with suspicion by the providers of distribution services, our analysis shows that there are clear regions in the parameter space where more distribution services are bought in equilibrium after the introduction of a P2P network. At the same time, this does not always happen: in other parameter regions, the market for content distribution services shrinks because unwanted demand shifts to the P2P network. When the P2P network has the potential to encroach on the publisher's demand, the optimal response of the publisher is to retaliate by increasing the speed (and sometimes the quality of its content as well). This, in turn, increases the spending on content provision. At the same time, we have also identified conditions under which the publisher finds it optimal to reduce investment in content provision services (both speed and quality) in response to the entry of P2P network and tolerate the lost demand.

With respect to publisher profits, the answer is again a qualified one: the introduction of a P2P network could either benefit or hurt the publisher's profits. The situation in which the publisher benefits is an interesting one: when excess (or unwanted) demand reaching the publisher is taken away by the P2P network, the publisher benefits. Under CPM pricing, a closer look reveals that this situation corresponds to a corner solution, i.e., when in the absence of a P2P network, the publisher chooses minimum (or normalized to zero) speed and quality. Because the publisher cannot reduce the speed and quality any further (because it would basically involve shutting down the website), there is no easy way of eliminating unwanted, low valuation consumers from accessing the content. If such consumers could be turned away from the site, it could benefit the publisher. Thus, in the CPM model, the benefit occurs from reducing content provision costs. In contrast, in the CPA model, the benefit could occur from increased advertising revenue because a higher per unit price per qualified action can be charged. In this scenario the P2P network acts as a filter because it self-selects low valuation (unwanted) consumers who prefer to consume the lower quality but speedier content from a P2P node. It is important to realize that the P2P entry leads to a pure cost-side effect (impact on content provision cost) in the CPM model whereas the effect could also be a demand-side one (impact on ad revenue) in the CPA model. Thus any increase in profit in the CPM model occurs from cost reduction, whereas in the CPA model, the profit may increase because ad revenue increases. As an extreme case, consider zero content provision costs. For this case, in the CPM model, the P2P entry can only hurt (at best, the profit is the same) the profit of the publisher. However, with zero content provision cost, the P2P entry can increase profit (i.e., the revenue) in the CPA model.

Advertising markets where ad revenue is accounted for using the CPA model may be more resilient to content piracy than those where the use of the CPM model is prevalent. This finding makes sense: in the CPM model the P2P's demand encroachment always results in lower revenue (never higher). However, in the CPA model the same level of demand encroachment could lead to more ad revenue. Thus the CPA model can be expected to tolerate content piracy better than the CPM model can.

The main conclusion from this study is that both publishers and content distributors alike need not always view the introduction of a P2P network as a threat. The manner in which advertising revenue is calculated has a strong influence on the incentives firms have to produce and distribute content. It should be noted that in the long run, P2P networks should not act to destroy the production of content because such networks cannot survive without original content. An aspect of the problem that may get interesting in the future is the commercial role P2P networks could play in the distribution of content. If P2P networks are to survive as a viable means of distributing content, then the commercial interests that drive the creation and distribution of content (namely, advertising and online commerce) cannot be separated from the efforts being made to advance P2P technology. For example, it may be possible for a publisher to legally license its content to the P2P network and share the revenue that the P2P network earns through its own advertising programs. Alternatively, the publisher may choose to manage its own distribution of content, perhaps using a combination of CDN technologies (e.g., for premium content) and P2P mechanisms (e.g., for standard content). In our model, we assume that content quality includes several aspects of the consumption experience such as currency, accessibility, security (extent to which access to information is appropriately restricted), and credibility (extent to which information is regarded as true or reliable). In future research, one could consider a more detailed model that separately considers these costs while determining the optimal content provision strategy for the publisher, e.g., the cost to the publisher to remove malware and guard against other security threats. Finally, the ideas in this study could be applied in a competitive setting, where two or more content providers that are competing for similar consumers have to consider the entry of a P2P pirate.

Electronic Companion

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/isre.1110.0406.

References

Bhatacharjee, S., D. R. Gopal, K. Lertwachara, R. M. Marsden, R. Telang. 2007. The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts. *Management Sci.* **53**(9) 1359–1374.

Damiani, E., D. C. D. Vimercati, S. Paraboschi, P. Samarati, F. Violante. 2002. A reputation-based approach for choosing reliable resources in peer-to-peer networks. *Proc. 9th ACM Conf. Comput. Commun. Security* (CCS '02J), ACM, 207–216.

Eisenmann, T. R. 2002. Akamai technologies. Harvard Business School Case no. 804–158. Harvard Business School Press, Boston.

Ercetin, O., L. Tassiulas. 2005. Pricing strategies for differentiated services content delivery networks. *Comput. Networks* **49**(6) 840–855.

Gallaugher, J. M., P. Auger, B. Anat. 2001. Revenue streams and digital content providers: An empirical investigation. *Inform. Management* **38**(7) 473–485.

Gopal, R. D., A. Gupta. 2010. Trading higher software piracy for higher profits: The case of phantom piracy. *Management Sci.* **56**(11) 1946–1962.

Hosanagar, K., R. Krishnan, M. Smith, J. Chuang. 2008. Service adoption and pricing of content delivery network (CDN) services. *Management Sci.* **54**(9) 1579–1593.

IAB Internet Advertising Revenue Report. 2011. Accessed April, 2011, http://www.pwc.com/us/en/industry/entertainment-media/publications/iab-internet-advertising.jhtml.

Jacob, V., K. Asdemir, N. Kumar. 2010. Pricing models for online advertising: CPM versus CPC. *Inform. Systems Res.* Forthcoming.

Julian, B. 2001. Business uses of peer to peer (P2P) technologies. Whitepaper, Netmarkets Europe. http://www.voidstar.com/downloads/PB2B.pdf

Knight, S., J. Burn. 2005. Developing a framework for assessing information quality on the World Wide Web. *Inform. Sci. J.* **8** 162–172.

O'Donnell, S. 2002. An economic map of the Internet. Research Paper 162, MIT Sloan School of Management, Cambridge, MA.

Oh, J. 2007. Profit sharing in a closed content market. *J. Media Econom.* **20**(7) 55–75.

Padmanabhan, V. N., P. A. Chou, H. J. Wang. 2003. Resilient peer-to-peer streaming. Technical Report MSR-TR-2003-11, Microsoft Research, Redmond, WA.

Pallis, G., A. Vakali. 2006. Insight and perspectives for content delivery newtorks. *Comm. ACM* **49**(1) 101–106.

Pan, J., T. Y. Hou, B. Li. 2003. An overview of DNS-based server selections in content distribution networks. *Internat. J. Comput. Telecommunications Networking* **43**(6) 695–711.

Parmeswaran, M., A. Susarla, A. B. Whinston. 2001. P2P networking: An information sharing alternative. *Comput.* **34**(7) 31–38.

Pipino, L. L., Y. W. Lee, Y. R. Wang. 2004. Data quality assessment. *Comm. ACM* **45**(4) 211–218.

Stolarz, D. 2001. Peer-to-peer streaming media delivery. *Proc.* 1*st Internat. Conf. Peer-to-Peer Comput.* (*IEEE*), Linkoping, Sweden, 48–52.

Vakali, A., G. Pallis. 2003. Content delivery networks: Status and trends. *IEEE Internet Comput.* **7**(6) 68–74.

Wallach, S. D. 2002. A survey of peer-to-peer security issues. Proc. 2002 Mext-NSF-JSPS Internat.Conf.Soft. Security: Theories and Systelms, Tokyo.

Yang, Z., S. Cai, Z. Zhou, N. Zhou. 2005. Development and validation of an instrument to measure user perceived service quality of information presenting web portals. *Inform. Management* **42** 575–589.