Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2003 Proceedings

International Conference on Information Systems (ICIS)

December 2003

An Economic and Operational Analysis of the Market for Content Distribution Services

Cuneyd Kaya
University of Texas at Dallas

Kutsal Dogan
University of Texas at Dallas

Vijay Mookerjee
University of Texas at Dallas

AN ECONOMIC AND OPERATIONAL ANALYSIS OF THE MARKET FOR CONTENT DISTRIBUTION SERVICES

Cuneyd Kaya, Kutsal Dogan, and Vijay Mookerjee
School of Management
University of Texas at Dallas
Richardson, TX USA
cckaya@utdallas.edu kutsal.dogan@utdallas.edu vjaym@utdallas.edu

Abstract

We develop an economic and operational model to examine the conditions for the viable provision of content distribution services by a monopolistic firm. Each user firm (the content provider or CP) has the option of buying content distribution services from the content distribution service provider (CDP) or going on its own to arrange for its content to be distributed at a set of chosen sites operated by Internet Service Providers (ISPs). The CDP enjoys operational benefits in terms of both the fixed and variable cost of replicating content. However, we find that for certain market situations (concentrated CP and ISP demand), not all CPs will find it attractive to buy services from the CDP. The best case for the CDP is when the various content providers have similar demand that is uniformly distributed across ISP sites.

Keywords: Content distribution, content providers, pricing of services

Introduction

The widespread growth and availability of the Internet has made it possible to deploy many new applications that can be accessed over a network. However, this unprecedented growth has a downside as well. As the traffic carried over the Internet increases, the performance of network-based applications often becomes unpredictably slow. A variety of techniques are being considered to improve the performance of networked applications including better capacity planning, class differentiation and reservation, and content distribution. This study is concerned with content distribution networks (CDN) as a technique to improve the performance of networked applications. While content distribution techniques apply to any networked application, we focus here on networked applications that, for the most part, use the public Internet as the means of connecting the various components necessary to make the application function.

The main idea behind content distribution is avoid congested network paths from having to carry client-server communications. This is achieved by providing a surrogate server site with which the client has better connectivity in terms of delay. Since a surrogate server will typically have better connectivity with a relatively small fraction of the clients, several surrogates will have to be provided to ensure acceptable performance for all client locations. Often content distribution services are obtained through a CDN provider, a firm that has special arrangements with numerous Internet Service Providers (ISPs) where surrogate servers may be located. Alternative distribution systems include peer-to-peer file sharing systems such as the Gnutella project (**http://gnutella.wego.com**; Kan 2001) and Kazaa (**www.kazaa.com**). Freenet (**http://freenet.sourceforge.net**) is a peer-to-peer file-storage system where the users can place their files on other nodes' storage spaces (Clarke et al. 2002; Langley 2001). However, large commercial content providers currently do not utilize peer-to-peer networking with other content providers for variety of reasons including competition and security.

The goal of this paper is to derive an economic and operational model to evaluate two options a firm may have to distribute its content: (1) by obtaining the services of CDN provider, and (2) by directly forming arrangements with the ISPs. We will hereafter refer to the firm that seeks to distribute its content as the *content provider* (CP), and the content distribution network provider as CDP. Current players in the content distribution market include U.S.-based companies Akamai (**www.akamai.com**) and Speedera (**www.speedera.com**), UK-based Cable and Wireless (**www.cw.com**), and Accelia (**www.accelia.net**) from Japan. CDPs act as intermediaries between content providers and ISPs. Intermediary services have been studied in the context of electronic marketplaces, where intermediaries are shown to enjoy aggregation benefits (Bhargava et al. 2000). Our model also considers benefits similar to aggregation that the CDP enjoys. However, we develop these benefits from an operational economy that the CDP derives, rather than benefits that accrue from aggregation of information.

Our work is related to studies on network pricing, information intermediary services, and allocation and pricing models of content delivery services. Existing studies on network pricing include pricing of congestible network resources such as an FTP server where pricing is used to ensure efficient use of these resources (MacKie-Mason and Varian 1995), priority-based pricing (Gupta et al. 1996; Mendelson and Whang 1990), optimal pricing of integrated services networks with a guaranteed quality of service (Wang et al. 1996), and a simulation of pricing policies in multiple service class networks (Cocchi et al. 1993). Corbett and Karmarkar (1999) provide an equilibrium pricing model for an information intermediary. Besides these pricing models, the Thomas et al. (2002) model develops an approach to allocate network resources optimally. In terms of pricing of content delivery systems, earlier research focuses on how content providers set the price they charge to end users. Jagannathan and Almeroth (2001) propose a probabilistic adaptive flat-fee pricing scheme for content delivery services based on user behavior. An extension to this study compares fixed and variable pricing schemes for delivering batched content such as video-on-demand and downloadable-CD. (Jagannathan et al. 2002). While prior research separately deals with pricing and resource allocation, our model jointly makes decisions on these two issues.

The content provider's objective is to ensure that many users see its content with acceptable delay. Often, ISPs cache some content, helping the content provider deliver its content closer to the users that demand the content. Hosanagar et al. (2002) develop a model for resource allocation and pricing for different levels of caching in ISP networks. A price-driven market-based resource allocation model for content delivery is studied by Ercetin (2002) using a noncooperative game theoretic approach among the content providers and surrogate servers.

At this stage of this research, we have been able to formulate a simple model and derive some qualitative results. These deal mainly with the link between the pattern of user demand for content and the viability of CDN service provision. Our main conclusion is that CDN services make sense when the CP side of the market is fragmented (no one CP dominates in terms of demand) and the user side of the market is also fragmented (no particular user location or ISP site dominates in terms of demand). As the demand structure moves away from the above, the CDP's position in the market weakens.

The Model

The model attempts to describe the market conditions that are viable for content delivery providers. CDPs offer content distribution services to content providers with a guaranteed level of service, measured in terms of average end-user delay. They accomplish this by putting surrogate servers within ISP locations. When a CDP puts content at an ISP location, the requests originating from that location do not have to travel on the public Internet. Hence users experience less delay. In addition, from the perspective of the ISP, there are benefits as well since the ISP can save a significant portion of the upstream bandwidth requirement. Achieving delivery of content this way is different from the caching done by ISPs. Typically, caching at an ISP is done from the perspective of the ISP, so only content that is in heavy demand is cached by the ISP. Second, while ISPs can cache some *content* on their own, a CDP replicates both content and *applications* on surrogate (or *edge*) servers at ISP sites.

The model envisions a monopolistic scenario with a single CDP serving many content providers (see Figure 1). There are n content providers (denoted by $CP_1, CP_2, ..., CP_n$) and m ISPs (denoted by $ISP_1, ISP_2, ..., ISP_m$) in the market. Each content provider is assumed to have a combination of static content, dynamic content, and applications that cater to end users in different geographical regions that access this content via ISPs. Note that an individual ISP in Figure 1 represents the site of an ISP rather than the entire ISP firm. Thus a combination of sites can be owned by the same ISP firm (such as America Online).

The key structure of our model is as follows. Content providers need an average delay guarantee and attractive price from the CDP in order to join to the CDN service. If the sum of the cost of the guaranteed delay and the CDP's price does not meet a content provider's *threshold cost*, the content provider should directly make arrangements with individual ISPs to locate surrogate servers at one or more ISP sites. The selection of ISP sites by a content provider will depend on the demand for the CP's content at specific sites. The aforementioned *threshold cost* is based on this selection and is used as a benchmark in the content provider's decision of whether or not to buy services from the CDP.

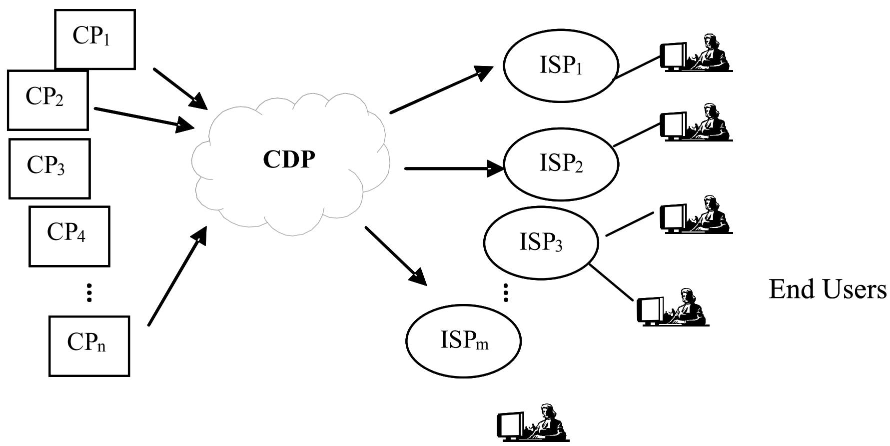

Figure 1. Architecture of a Content Delivery Network

The CDP distributes content at a number of ISP sites that are again selected based on the demand for content across the participating CPs. However, the CDP has two operational benefits. The first benefit is the fixed cost of opening a new ISP site. The CDP pays a fixed cost for the first time a surrogate server is placed at a new ISP site. After an ISP site has been opened, there is only an incremental cost for adding new content. This incremental cost is the cost of keeping the content current. Here too, the CDP enjoys scale economies.

The mechanism works as follows. First, each CP solves a cost minimization problem to decide on the subset of ISP sites at which the content will be replicated. The CDP's solution will have to beat the CP's solution, otherwise the CP will not buy CDN services and will choose to directly deal with the ISPs. The CDP will obtain its profit maximizing solution by considering all the CPs and offer CDN services to each CP at a price and an average delay guarantee.

CP's Problem

Let **D** denote the demand matrix where d_{ij} represents the demand for the content provider *i* (CP_i) at j^{th} ISP (ISP_j). Let **L** denote the expected delay matrix of delays among different ISPs where L_{ij} represents the amount of delay between i^{th} and j^{th} ISPs (ISP_i and ISP_j). Given these two matrices, i^{th} content provider (CP_i) solves a cost minimization problem that is described below.

The objective is to minimize the sum of delay cost and fixed and variable cost of replication by choosing a set of sites at which to replicate the content (x_{ij}s). For any given set of assignments to the x_{kj} variables, the average delay for a CP is calculated as follows. The delay is zero for a demand at an ISP that carries the content. At an ISP site that does carry the CP's content, the delay is calculated by obtaining the content from the ISP site that has the least inter-ISP delay. The CP's problem is formally expressed below.

$$\text{Min } \; TC_{CPi} = C(\delta_i) + \sum_{j=1}^{m} x_{ij} F + R(\sum_{j=1}^{m} x_{ij})$$

Subject to

$$x_{ij} \in \{0,1\}$$

Where $\delta_i = \dfrac{\sum_{j=1}^{m} d_{ij}(1-x_{ij}) \underset{l \neq j, x_{il} \neq 0}{Min} \{L_{jl}\}}{\sum_{j=1}^{m} d_{ij}}$ and *l=1, 2, …, m.*

The content provider's decision variables (x_{ij}) take binary values as follows:

$$x_{ij} = \begin{cases} 1, & CP_i \text{ replicates its content on } ISP_j \text{'s network,} \\ 0, & \text{Otherwise.} \end{cases}$$

The solution of this problem will give CP_i an optimal cost-delay pair (TC_{CP_i} , δ_i). The fixed cost parameter F represents the cost of placing the content at any ISP location. The replication cost R is modeled as a concave function of the total number of replications. As this number increases, the replication cost increases at a decreasing rate. This is because of economies in propagating updates to surrogate sites that are relatively close to one another. Previous studies have shown that delay costs (C(δ)) are often nonlinear (Ercetin 2002). Note that, in the content provider's problem, demand is expressed as data demanded per period. Similarly, all of the costs are operating costs that are expressed as dollar amounts per period.

CDP's Problem

CDPs typically price content providers that join to their network according to a two-part tariff scheme that includes a fixed charge plus a usage fee based on the amount of traffic users of that CP generate on the surrogate servers. The CDP also provides the CP with an average delay guarantee. The CDP must determine the price p_i it will charge to each content provider given by $p_i = \alpha + \sum_{j=1}^{m} \beta d_{ij}$, where α is the flat portion of the price and β is price per Mbps of user demand for the CP's content at an ISP. The delay guarantee for a given CP can be calculated using the CDP's replication variables for that CP and the delays that would be incurred at each ISP site that does not carry the content.

The objective of the CDP is to maximize profit. The total revenue (TR) is the sum of the revenue collected from each CP that buys its content distribution services. The total cost (TC) is the total fixed cost of content distribution (one fixed cost per ISP site that is used for replication) plus the total replication cost that arises from the need to keep the content current. Note that the CDP can be expected to enjoy greater geographical economies than the CP since there will be more content replication (across CPs) in the CDP's solution. The CDP must choose the price variables (α and β) the replication variables (y_{ij}), and the participation variables (q_i) to maximize profit. The CDP solves the following problem.

$$\text{Max } \Pi_{CDN} = TR - TC = \sum_{i=1}^{n} q_i(\alpha + \beta \sum_{j=1}^{m} d_{ij}) - F \sum_{j=1}^{m} (I(\sum_{i=1}^{n} y_{ij})) - R(\sum_{j=1}^{m} \sum_{i=1}^{n} y_{ij})$$

Subject to

$$TC_{CP_i} \geq q_i \left(c(\delta_i) + \left(\alpha + \beta \sum_{j=1}^{m} d_{ij} \right) \right) \text{for all } i = 1, 2, \cdots, n$$

$$q_i \leq \sum_{j=1}^{m} y_{ij} \quad \text{for all } i = 1, 2, \cdots, n$$

$$y_{ij} \in \{0,1\} \text{ and } \alpha, \beta \geq 0 .$$

Where, $\delta_i = \dfrac{\sum_{j=1}^{m} d_{ij}(1-y_{ij}) \underset{l \neq j, y_{il} \neq 0}{Min} \{L_{jl}\}}{\sum_{j=1}^{m} d_{ij}}$ and $I(t) = \begin{cases} 1, & t \geq 1 \\ 0, & t < 1 \end{cases}$ is the indicator function that calculated as

$$I(\sum_{i=1}^{n} y_{ij}) = 1 - \prod_{i=1}^{n}(1-y_{ij}).$$

The CDN provider's decision variables (y_{ij} and q_i) take binary values as follows:

$$y_{ij} = \begin{cases} 1, & \text{Content of } i^{th} \text{ content provider is replicated on } ISP_j\text{'s network} \\ 0, & \text{Otherwise,} \end{cases}$$

$$q_i = \begin{cases} 1, & \text{if } CP_i \text{ joins the CDP's network,} \\ 0, & \text{Otherwise.} \end{cases}$$

Then for each content provider CP_i, the CDP will have the total expected delay for its content, $\delta(y_{ij})$.

As in the case of the CP, *F* represents the fixed cost of opening a new ISP site for the CDP. Here, the CDP benefits from scale economies since in the CDP's solution, content from different CP's can be located at the same ISP site. Note that this fixed cost economy does not apply to the CP since the same content will never be duplicated at the given ISP location. Furthermore the cost of keeping the updates current will also be less for the CDP (on a per content basis) because of geographical economies. Similar to the content provider's problem, all costs and demand for content are expressed as per period quantities. After solving its problem, the CDN provider will offer content distribution services to each CP at a given price and delay guarantee (p_i, δ_i) that will allow each content provider to decide whether or not to buy CDN services.

Solution Procedure

The preceding two problems are hard problems (*NP*-complete) (Bussieck and Pruessner 2003). The content provider's problem is a nonlinear integer programming (INLP) problem whereas the CDP's problem is a Mixed Nonlinear Integer Programming (MINLP) problem. For large values of *m* and *n,* these two problems are computationally expensive to solve. In the content delivery market there are approximately 7 million content providers (200,000 of them being commercial content providers) and thousands of ISP sites (Eisenmann 2002); therefore, a CDP's problem can possibly have millions of decision variables. Of course, firms can use preprocessing to eliminate certain content providers and ISP sites. Nevertheless solving the two problems will not be trivial. Work is underway to find better solution techniques (Bussieck and Pruessner 2003).

Example

To shed light on the inner dynamics of the model, consider a small example with eight content providers *(n = 8)* and four ISPs *(m = 4)*. To ensure tractability, simple but representative expressions are used for the delay cost and the replication cost functions. For the delay cost function $C(\delta_i) = c\delta_i^2$ is used where *c* is a delay cost parameter. For the replication cost $R(r) = g\left(1 - \frac{1}{\sqrt{1+r}}\right)$ is used where *g* is a replication cost parameter and *r* is the number of replications, given by $r = \sum_{j=1}^{m} x_{ij}$ for the i^{th} content provider. We have set *c* = *g* =1 for simplicity, in this version of the paper.

The expected delay (per MB) matrix $\mathbf{L} = \begin{pmatrix} M & 2 & 4 & 4 \\ 2 & M & 4 & 4 \\ 4 & 4 & M & 8 \\ 4 & 4 & 8 & M \end{pmatrix}$ is designed to incorporate two different types of ISPs: small and large. The first two ISPs are large and the connectivity between the two is very good, leading to low expected delay. The third and fourth ISPs are small so the connectivity between the two is poor. Additionally the expected delay from the two large ISPs to the two small ones is better than the connectivity between the small ones but worse than the connectivity between the two large ones. To facilitate the use of **L** in calculation of average delays and allow the use of *Min*(•) function, the diagonal values are set to M, which is a sufficiently large number. Four cases, based on different demand characteristics (see Table 1), are described and implications for the content providers and the CDN provider are discussed.

Table 1. Demand Structure

	Uniform ISP Demand	*Concentrated ISP Demand*
Uniform CP Demand	**Case 1**	**Case 4**
Concentrated CP Demand	**Case 2**	**Case 3**

Case 1

$$\mathbf{D} = \begin{pmatrix} 8 & \dots & 8 \\ \vdots & \ddots & \vdots \\ 8 & \cdots & 8 \end{pmatrix}$$

In this case, the demand for content is uniformly distributed among *n* content providers. Furthermore, each content provider's demand is also uniformly distributed across the *m* ISPs. Because of this uniformity, content providers behave similarly; therefore, the decision is simple for both the CDP and the content providers. The decision for content providers is based on the values of the fixed cost parameter *F* and the delay cost parameter *c*. For a sufficiently small value of *F* (large *c*) content providers should place their content at all of the ISPs. As *F* increases (*c* decreases) ,the decision to place content at all ISPs will change to placement at a subset of ISPs. Finally for a sufficiently high large *F* (small *c*), the content will be placed at a single ISP. Note that the content has to be placed with at least one ISP to avoid infinitely large delays. This (single) ISP can be thought of as the content provider's origin site. For the CDN provider, the decision is also quite straightforward. The price parameters (α and β) can be chosen so as to attract all content providers to the CDN by offering them a slightly cheaper alternative. This case is the best one for the CDP.

Case 2

$$\mathbf{D} = \begin{pmatrix} 32 & 32 & 32 & 32 \\ 32 & 32 & 32 & 32 \\ 32 & 32 & 32 & 32 \\ 32 & 32 & 32 & 32 \\ 8 & 8 & 8 & 8 \\ 8 & 8 & 8 & 8 \\ 8 & 8 & 8 & 8 \\ 8 & 8 & 8 & 8 \end{pmatrix}$$

The demand matrix for this case is shown above. In one aspect,this case is quite similar to case 1. The similarity stems from the resulting average delay. Note that as long as the distribution of the demand for two content providers is uniform across the ISPs, the average delay will be the same for these content providers. However, in another aspect, case 2 is quite different from case 1. In case 2, the CDP has less flexibility in the pricing decision; hence there is loss of market power. The tradeoff between the two price parameters (α and β) causes the relative lack of power. If the CDP sets a high value for α and a low value for β, small content providers are penalized and these providers may not join the network. Alternatively, if the CDN provider sets a low value for α and a high value for β, large content providers are penalized and these providers may not join the network. In either case, the CDP loses a certain portion of its profits.

Case 3

Case 3 depicts a demand pattern that is concentrated at certain ISPs and also concentrated toward certain content providers. In this case, the optimal solution to the content providers' problem shows that for sufficiently high F values, content providers with concentrated demand will have less incentive to join to the CDN. The CDP, however, is able to attract content providers with unconcentrated demand. For small values of F, the CDN provider can attract all content providers to join the CDN network.

Case 4

$$\mathbf{D} = \begin{pmatrix} 32 & 32 & 8 & 8 \\ 32 & 32 & 8 & 8 \\ 32 & 32 & 8 & 8 \\ 32 & 32 & 8 & 8 \\ 32 & 32 & 8 & 8 \\ 32 & 32 & 8 & 8 \\ 32 & 32 & 8 & 8 \\ 32 & 32 & 8 & 8 \end{pmatrix}$$

This case is the opposite of case 2 with a demand pattern that is concentrated at certain ISPs but uniform across the content providers. For very small F values, content providers should replicate content on all of the ISPs. As F increases, content providers should favor ISPs with large demand. Depending on the value of F, the CDN can attract content providers to join the network by adjusting the price parameters(α and β). Similar to case 2, the CDP's market power is less in comparison to the uniform-uniform case. The loss of power occurs because content providers can focus their content on those ISPs that have large demand for the content. This is a relatively low cost solution that the CDP may find difficult to beat.

Limitations and Future Work

Our model is limited in that it assumes that an ISP can handle any amount of content that is put on its servers, i.e., ISP capacity is not an issue. Even though the capacity in terms of storage may not be a serious issue for many ISPs, capacity in terms of bandwidth should be considered. This issue is currently being addressed in an extended model. Another limitation of the model has to do with the assumption of deterministic demand that does not change with delay. This can be incorporated in the model by letting demand matrix **D** depend on the delay. The issue of linking demand with delay has broader implications as well. There is no explicit modeling of user behavior or ISP behavior in the current model. In reality, an ISP also benefits if delays are reduced and may therefore agree to subsidize certain CDP (or large CPs) in return for interesting or popular content. We believe that CDPs will exert greater power in such situations since they act as aggregators of content. This issue currently is being addressed to develop a more complete model of the players in the content distribution chain and their incentives. This chain consists of CPs (who want many end users to access their content), CDPs (who exploit scale economies of distribution and content aggregation), ISPs (who strive to reduce user delay while keeping their network and storage costs under control), and end users (who pay a price for ISP services and have a disutility for access delay).

References

Bhargava, H. K., Choudhary, V., and Krishnan, R. "Pricing and Product Design: Intermediary Strategies in an Electronic Market," *International Journal of Electronic Commerce* (5:1), 2000, pp. 37-56.

Bussieck, M. R., and Pruessner, A. "Mixed-Integer Nonlinear Programming," GAMS Development Corporation, Washington, DC, February 19, 2003 (available online at **http://www.gamsworld.org/minlp/siagopt.pdf**; accessed May 1, 2003).

Clarke, I., Miller, S. G., Hong, T. W., Sandberg, O., and Wiley, B. "Protecting Free Expression Online with Freenet," *IEEE Internet Computing* (6:1), January-February, 2002, pp. 40-49.

Cocchi, R., Shenker, S., Estrin, D., and Zhang, L. "Pricing in Computer Networks: Motivation, Formulation, and Example," *IEEE/ACM Transactions on Networking* (1:6), December 1993, pp. 614-627.

Corbett, C. J., and Karmarkar, U. S. "Optimal Pricing Strategies for an Information Intermediary," Decisions, Operations and Technology Management Working Paper-CORB003, The Anderson School, University of California, Los Angeles, August 29, 1999.

Eisenmann, T. "Akamai Technologies," Harvard Business School Case 9-802-132, March 2002.

Ercetin, O. *Market-Based Resource Allocation for Content Delivery in the Internet*, Unpublished Ph.D. Dissertation, University of Maryland, College Park, 2002.

Gupta, A., Stahl, D. O., and Whinston, A. B. "Priority Pricing of Integrated Services Networks," in *Internet Economics*, L. W. McKnight and J. P. Bailey (eds.), The MIT Press, Cambridge, MA, 1996, pp. 323-352.

Hosanagar, K., Krishnan, R., Chuang, J., and Choudhary, V. "Pricing and Resource Allocation in Caching Services with Multiple Levels of QoS," in *Proceedings of Twenty-Second International Conference on Information Systems,* L. Applegate, R. Galliers, and J. I. DeGross (eds.), Barcelona, Spain, December 2002, pp. 481-491.

Jagannathan, S., and Almeroth, K. C. "An Adaptive Pricing Scheme for Content Delivery Systems," in *Proceedings of IEEE Conference on Global Commerce*, Volume 3, San Antonio, TX, November 2001, pp. 1641-1645

Jagannathan, S., Nayak, J., Almeroth, K., and Hofmann, M. "On Pricing Algorithms for Batched Content Delivery Systems," *Electronic Commerce Research and Applications* (1:3/4), 2002, pp. 264-280.

Kan, G. "Gnutella," in *Peer-to-Peer: Harnessing the Benefits of a Disruptive Technology*, A.Oram (ed.), O'Reilly & Associates, Sebastopol, CA, 2001, pp. 94-122.

Langley, A. "Freenet," in *Peer-to-Peer: Harnessing the Benefits of a Disruptive Technology*, A.Oram (ed.), O'Reilly & Associates, Sebastopol, CA, 2001, pp. 123-132.

MacKie-Mason, J. K., and Varian, H. R. "Pricing Congestible Network Resources," IEEE *Journal on Selected Areas in Communications* (13:7), September 1995, pp. 1141-1149.

Mendelson, H., and Whang, S. "Optimal Incentive-Compatible Priority Pricing for the M/M/1 Queue," *Operations Research* (38:5), September-October 1990, pp. 870-883.

Wang, Q., Peha, J. M., and Sirbu, M. A. "Optimal Pricing for Integrated Services Networks," in *Internet Economics*, L. W. McKnight and J. P. Bailey (eds.), The MIT Press, Cambridge, MA, 1996, pp. 353-376.

Thomas P., Teneketzis, D., and MacKie-Mason, J. K. "A Market-Based Approach to Optimal Resource Allocation in Integrated-Sevices Connection-Oriented Networks," *Operations Research* (50:4), July-August 2002, pp. 603-616.