Network pricing: can both ISP and P2P benefit?

Zhen Li[1] and Qi Liao[2],*,†

[1]Department of Economics and Management, Albion College, Albion, MI, USA
[2]Department of Computer Science, Central Michigan University, Mount Pleasant, MI, USA

SUMMARY

Internet service providers are facing increasing back pressure from rising access demand by users, especially peer-to-peer (P2P)-based applications that greatly enhance the large-scale distribution of content into and out of their networks. With the ever increasing consumption pressure on scarce bandwidth resources, ISPs have been forced to reconsider their business model of overselling 'all-you-can-eat' broadband at flat rates. Technical solutions such as traffic differentiation or blocking violate the principle of network neutrality; traffic shaping and deep packet analysis fall short in the presence of encryption; and P4P (localized P2P)-based solutions are difficult to achieve in a heterogeneous environment. Economically, various usage-based pricing schemes have been proposed and discussed. While they can improve efficiency in bandwidth consumption, they tend to face strong customer resistance as users have strong preference in favour of simple flat rates. We argue that any feasible pricing reforms cannot deviate much from the current flat rates while providing financial incentives for bandwidth hogs to limit their bandwidth access. In contrast to normal usage-based pricing models that charge by volume, we propose a temporal-based pricing model that may generate a mutually beneficial solution that can not only increase the profitability of ISPs but also accommodate P2P, rather than killing it, without changing the software, protocols or hardware that clients or ISPs use on the network. Copyright © 2014 John Wiley & Sons, Ltd

1. INTRODUCTION

Internet service providers offer users access to the Internet and related services. There are roughly two tiers of ISPs according to the range of services provided: *local* ISPs that provide services in small regions to end-users and *transit* ISPs that transfer data between local ISPs [1]. Local ISPs are essentially last-mile access providers, who sell Internet access to end-users. Conventionally, by paying a flat monthly rate, users are entitled unlimited access like an 'all-you-can-eat' bandwidth buffet. Nevertheless, the bandwidth resources of ISPs are not unlimited. Just as their customers pay them for Internet access, local ISPs themselves pay upstream ISPs for Internet access. As bandwidth consumption keeps growing, the practice of limiting or charging more for large bandwidth usage has been a hot subject in the tech community in recent years.

The soaring demand for bandwidth is driven by lots of applications, noticeably peer-to-peer (P2P) file sharing. While the traditional client–server, unicast model is still the dominant form of content distribution on the Internet, P2P architectures [2–4] have become popular and sprouted up to provide an alternative way of transferring large amount of data by utilizing the diverse connectivity available among participants in a network and the cumulative bandwidth of its participants. In recent years, although easily accessible streaming media such as Netflix and YouTube have led to a downward trend in P2P file sharing of movie and music, P2P systems such as BitTorrent still dominate upstream

---

*Correspondence to: Qi Liao, Department of Computer Science, College of Science and Technologies, Central Michigan University, Mount Pleasant, MI 48859, USA.
†E-mail: liao1q@cmich.edu

traffic with 36.35% as of 2013 [5]. In addition, P2P still has its uses for legitimate file distribution of large content. With the motivation to scale and the appealing feature of utilizing untapped end-user bandwidth and CPU/disc resources, P2P has received significant attention so as to harness the technology as a mechanism for legal content distribution [6–8]. Some recent findings show signs of the resurgence of P2P while moving P2P service to the cloud via the notion of a 'seedbox' [9].

Unfortunately, P2P applications put a strain on the infrastructure of ISPs. Compared with other applications (e.g. Web and email traffic) that have clear on/off patterns, P2P applications by nature serve as both client and server and tend to have a longer duration. They may require equal uplinks as well as downlinks and are likely to become bandwidth hogs that consume a larger share of the transmission capacity than traditional applications. By paying the same monthly flat rate, heavy P2P users are subsidized by normal users for their unfair share of bandwidth consumption. Because of the increased load and changing traffic patterns resulting from the large amount of traffic generated [6], especially the upstream bandwidth requirements by P2P, ISPs have been under significant pressure on their maximum link capacity to domains outside of their local networks, that is, their bottleneck link capacity. ISPs' common actions to block or degrade certain applications (e.g. BitTorrent) [10,11] and attempts to differentiate traffic [12] are all signalling the bandwidth scarcity suffered by ISPs. It has also been shown that ISPs are often not profitable, as the increasing transit traffic leads to rising costs while Internet access prices have been falling, typically in the form of flat access rates [13].

Therefore, the conventional and dominant business model of ISPs that charges a flat monthly service fee for end-users to have unlimited access is not sustainable. It has been at tension against the challenge of widespread P2P practices accompanied by overall increase in access demand. As is the case for any type of unlimited services, flat rates provide no incentives for users to limit their consumption at zero marginal cost. Although the majority of the users will never reach the maximum usage capacity, a handful users will test the limits of the 'unlimited' access. According to a Time Warner Cable study, 5% of the company's broadband customers have been found to consume more than 50% of the total bandwidth [14]. Facing the soaring demand for bandwidth, ISPs are having trouble multiplexing their users as before because the patterns of usage are changing as a result of P2P applications. This leaves the ISPs with few choices: they are either forced to expand capacities, to throttle certain types of traffic, or to change their business model (including the introduction of bandwidth caps and tiered-usage pricing). In recent years, ISPs have begun shifting towards usage-based pricing that offers customers a fixed amount of data each month for a fee. On average, less than 2% of users exceed the most commonly used tier of 300 GB; nearly 80% of consumers never exceed even 50 GB per month [15].

The capacity pressure on ISPs caused by applications such as P2P can in turn interfere with the development of P2P. For instance, if P2P traffic were throttled or blocked or if usage-based pricing induced more P2P participants to become free riders, the availability and performance of P2P networks would deteriorate. We have observed past trends in P2P research to change the technical and/or economic incentives of users to deal with the free-riding problem [3,16–26]. The sustainability of both the ISP business model and P2P applications is in challenge. It seems like the prosperity of one may come at the cost of the other. Recent works on ISP-compliant P2P applications [27] and geographical-aware P4P [28] attempt to address the interest conflicts between ISPs and P2P applications in a technical way by changing how P2P clients connect with each other. While the bottleneck traffic jam associated with P2P can be mediated by techniques such as P4P, they do not help much when P4P trackers and groups on a network are heterogenous.

In addition to seeking technical solutions, ISPs have been also reconsidering their economic roles. *Can we find a pricing model as opposed to flat rates and usage-based pricing for ISPs that benefit both ISPs and users without requiring changes in the network protocols or applications used? Is it easy to implement a mutually beneficial pricing model without causing much consumer resistance?* In this paper, we seek for the answers to such questions by studying the impact of bandwidth pressure on ISPs. We propose a temporal-based pricing scheme that imposes surcharges beyond the data allowance on the basis of certain time windows, for example, on a daily basis. Temporal-based pricing combines the advantages of both unlimited bandwidth access (flat rates) and payment-induced consumption models (metered rates), providing a *mutually beneficial* solution for both Internet users and ISPs. Besides increasing ISPs' revenue and profitability, more importantly, the temporal-based pricing

scheme prevents the hike of the flat rate while keeping the quality of service (QoS) for end-users. It may even reduce the churn rate of P2P, as opposed to usage-based pricing; thus, it is more P2P friendly.

The remainder of this paper is organized as follows. Section 2 reviews the existing research in the literature. Section 3 discusses the flat-rate business model of ISPs and analyses how rising bandwidth demand caused by P2P and other applications may threaten the profitability of ISPs. Section 4 introduces a temporal-based pricing model, evaluates the interactions between ISPs and P2P applications, and shows that temporal-based pricing can be mutually beneficial. Section 5 discusses the benefits of the temporal-based pricing scheme and studies a number of key implementation issues for shifting from the conventional flat rates to temporal-based pricing. Section 6 concludes with summarizing remarks.

2. RELATED WORK

Much of the past P2P-related research has been focused on incentive mechanisms to prevent free riding. The proposed solutions to motivate P2P participation is typically to provide either technical or economic incentives. Technical approaches associate bandwidth allocation of downloading with the corresponding uploading contribution such as the 'tit-for-tat' in BitTorrent-like systems [19,29]. As a supplement to the exchange-based incentives, payment-based incentive mechanisms [3,17,18] attach a real monetary cost to low availability. Other incentive mechanisms include optimal resource alloca- tion and admission control ranging from rank-based [22,25] to reputation-based mechanisms such as EigenTrust [30]. Other proposed schemes include distributed incentive schemes such as MARCH [26] and fairness-enforcing infrastructures such as Samsara [31].

While incentives to eliminate free riding in content-sharing applications are important, researchers should not neglect the bandwidth pressure the increased capacity demand may impose on ISPs' limited resources because the business model of ISPs is a key factor affecting the fate of P2P, that is, the avail- ability of resources that are not really there for free. We are motivated by the missing link between ISP profitability and the prosperity of P2P applications and look for possible mutually beneficial solutions that can foster both ISPs and P2P. That is, what if P2P applications successfully discourage free riding and more nodes are contributing? What are the impacts of rising P2P demand for bandwidth capac- ity on current ISPs' bandwidth subscription and pricing models? Many questions remain unanswered regarding the conflicts between ISPs' scarce capacity resources and increased demand for resources by heavy usage of P2P and other applications.

There has been research about the rising tension between P2P and ISPs. Peer Coordination Protocol [27] was developed for ISP-compliant P2P systems, which puts a dynamic rate limit on P2P traffic based on an ISPs' constraint. Application-layer traffic optimization studies how to provide the right type of network layer topology information to the requesting peer to enable P2P data structure to improve performance while reducing the utilization of the underlying network infrastructure [32,33]. In [34], researchers conducted a simulation study to examine how localizing P2P traffic within network boundaries impacts the profitability of ISPs. They found that the benefits of localization must not be taken for granted. Wang *et al*. [35] modelled the peering and routing tussle between ISPs and P2P applications and analysed the economic implications of overlay traffic on ISPs' peering decisions. While those works focused on the peering relationship, we study the pricing model of ISPs for last- mile consumers. Not specifically targeting P2P networks, work in [36] studied the ISPs' equilibrium behaviour in an Internet QoS market with flat rates and usage-sensitive pricing schemes. The temporal- based pricing scheme we recommend is a combination of flat rates and usage-based pricing so as to maximize ISPs' profit and to increase P2P user surplus in the meantime under the current Internet best- effort model. In [6], authors pointed out the top 10 issues needing careful consideration before P2P can be used for legal content distribution. We carry out an economic analysis on two specific issues, P2P impact on ISPs and their business models and the proof of an optimal pricing scheme.

This paper seeks optimal ISP pricing schemes in the setting of rising tension between file- sharing applications and ISPs. Our work is closely related to a number of research focusing on the business models of bandwidth providers. Dhamdhere and Dovrolis [37] viewed the network neutrality debate as really about the profitability of ISPs and showed that certain strategies such as heavy-hitter

usage-based pricing are sensitive to user reaction. Wu *et al.* [38] studied the network congestion management by ISPs. Tran and Tuffin [39] focused on inter-domain routing and economic incentives through pricing strategies. Odlyzko [40] reviewed the pricing of telecommunications from a historical view. Geng *et al.* [41] and Nabipay *et al.* [42] studied bundling marketing strategies of information goods. All these works found flat rates to be optimal under certain conditions. Nevertheless, the capacity pressure faced by ISPs due to heavy users of P2P (and other applications) does not satisfy the conditions; thus, bundling pricing fails to be optimal. By avoiding strong confrontation with consumers' preference for simplicity, the proposed temporal-based model can lead to more effective use of ISPs' network and generate more revenue for ISPs. A dynamic, time-dependent usage pricing was proposed in [43], which charges users different prices for Internet access at different times. The purpose of the work is not to reduce demand by individual heavy users, but rather to prevent the peak demand by all users. It can be P2P unfriendly as the high charge on peak hours may cause high churn rates of P2P.

3. THE CONFLICTS BETWEEN ISPS' BUSINESS MODEL AND P2P

In this section, we first introduce the general business model of ISPs and then discuss why and how the rising demand for capacity caused by P2P and other applications may cause the business model of ISPs to be unsustainable.

3.1. The business model of ISPs

There are local ISPs that provide services in small regions to end-users versus transit ISPs that transfer data between local ISPs [1]. The ISPs of our focus are the local ISPs that sell the Internet capacity they acquire from upstream transit providers to subscribed end-users. While end-users pay the local ISPs, the local ISPs in turn pay the transit ISPs. ISPs earn their revenue mostly from end-users, and they incur costs to operate their network and to purchase upstream connectivity from transit providers. The common business model of the ISPs is to oversell Internet services while advertising unlimited access at a subscribed bandwidth at a flat monthly fee [36].

Let C be the bottleneck capacity the ISP has purchased from the upstream transit provider, which is the maximum Internet traffic the ISP can handle at any point in time without sacrificing the QoS. ISPs commonly provide asymmetric bandwidth subscription where the uplink is a small fraction of the downlink. Suppose the ISP currently has N users and a typical user i subscribes to b_i downlink bandwidth from the ISP. We define $(\alpha \cdot b_i)$ as the uplink bandwidth subscription by user i where $0 < \alpha < 1$.

Combining all N users, the total bandwidth sold by the ISP is $\sum_{i=1}^{N}(1 + \alpha)b_i$. If the ISP never oversells, that is, if $\sum_{i=1}^{N}(1 + \alpha)b_i \leq C$ always holds true, all users would be guaranteed the subscribed bandwidth all the time. Nevertheless, ISPs are known for overselling Internet bandwidth as $\sum_{i=1}^{N}(1+\alpha)b_i \gg C$, with the expectation that at any point in time, only a fraction of users may require Internet access and they are unlikely to saturate subscribed bandwidth either. If end-users behave as expected without saturating subscribed bandwidth most of the time, the business model of ISPs would work well.

Consider a basic profit function of the ISP. Given the number of end-users (N) and the monthly flat rate paid by each user $P(b_i)$ that is normally increasing in the size of bandwidth subscription, the ISP's total revenue from selling bandwidth access to users is $\sum_{i=1}^{N} P(b_i)$. The costs of the ISP include mainly the cost of buying bottleneck bandwidth capacity from the upstream provider and operating costs. The former is essentially a large-amount lump-sum cost incurred when the ISP expands bandwidth capacity [44]. For simplicity, we assume the ISP's costs depend completely on the bottleneck bandwidth capacity. The profit earned by the ISP is therefore

$$\pi = \sum_{i=1}^{N} P(b_i) - L(C) \tag{1}$$

where $L(C)$ is the ISP's lump-sum price of purchasing certain bottleneck capacity from the upstream transit provider. It increases with the bottleneck capacity.

At flat access rates paid by users, the revenue of the ISP does not depend on the actual traffic flow of the existing users. It increases only when the number of service subscribers increases. At a given monthly flat rate and the number of end-users, the revenue level of the ISP is fixed. As the traffic flow increases, the capacity pressure on the ISP increases, and so does the pressure on the ISP's profitability: the profitability of the ISP would stay the same until the rising bandwidth demand eventually forces the ISP to expand bottleneck capacity. As L increases at a larger capacity C, the profitability of the ISP would decrease.

As mentioned, the capacity expansion driven by the increased bandwidth demand of existing users is not profitable to the ISP, but very often, the ISP has to purchase more bottleneck capacity because for the ISP, there exists contingent down costs that would occur when the ISP is either down or losing customers to competing ISPs. The contingency depends upon the ISP's ability to maintain the QoS to satisfy users. QoS falls when the increased traffic flow jams the bottleneck capacity. There are often more than one ISP serving the same region. It has been observed that the bandwidth prices are roughly comparable within a geographic region, and for local ISPs, the competition is between how the demand is distributed across competing ISPs [45]. Hence, if the QoS of the ISP falls below a threshold that its users can no longer bear, the ISP will have to purchase more bandwidth access from its transit ISPs to release capacity pressure; otherwise, it will start losing customers to competing ISPs and eventually fall out of business [45].

3.2. P2P challenge on the ISP business model

Based on the business model of ISPs, overselling unlimited access at flat rates can be profitable if few users saturate their subscribed bandwidth. As shown in [42], when the marginal cost of serving extra demand is zero or very low, flat rates as a form of bundling are almost always more profitable to the monopoly service provider than separate sales, even when there are some buyers with disproportionately large usage. Nevertheless, flat rates offer no incentive for users to limit bandwidth consumption. Economic theory suggests that sunk costs do not affect marginal decision-making. For users, the fixed subscription cost is sunk, and the marginal cost of access is zero. The increasing penetration of broadband access, faster last-mile links and the rise of Internet video all mean that users are downloading increasingly more content.

Peer-to-peer applications make the capacity scarcity of ISPs more severe. In contrast to other traffic flows from content providers to ISPs, traffic flow generated by P2P applications can be both downlink and uplink. Because P2P users may frequently serve content to users outside their local ISPs, the use of P2P applications can cause a tremendous increase in the consumption of bandwidth available in the uplink. The extended duration of access occupying a larger share of subscribed bandwidth is putting even more pressure on ISPs' bottleneck bandwidth. P2P applications, when constituting a significant percentage of network traffic, use ISP bandwidth especially via uplinks without providing any specific compensation to ISPs (or other users). On one hand, ISPs are reluctant to expand bandwidth merely to accommodate P2P, which would only incur additional cost without generating revenue compensation, and thus reducing profit. On the other hand, if ISPs stay with the current bottleneck bandwidth, the heavy service demand by P2P will cause high congestion and may result in paying large down cost as users migrate or lawsuits occur and again reduce profit.

How should ISPs react to the threat caused by P2P applications? All in all, ISPs will continue to oversell bandwidth to avoid any capacity waste and to generate more revenue and market share from a large pool of customers. In response to P2P traffic, ISPs have the option of expanding capacity by upgrading their infrastructure and buying more bottleneck bandwidth (especially the uplink capacity). However, this is very costly to ISPs, and the resulting significant marginal cost will make flat rates fail to be optimal. What is more, the expanded bandwidth can be saturated again by P2P traffic that forces ISPs to expand further, which is analogous to how more highways are built only to be jammed by more traffic (due to both the increased number of cars and more intense driving per car).

As ISPs see their profitability shrink as their role in the Internet ecosystem becomes to 'move bits around' rather than to provide end-to-end services, they want to change the status quo by charging

some content providers, differentiating traffic in priority classes, and degrading or entirely blocking certain flows. In reality, ISPs are reluctant to provide extra bandwidth to accommodate P2P because the fall in profit would be fearsome. Instead, they have been trying various practices to be profitable while they continue to charge their customers a flat price. For example, ISPs prefer to prioritize traffic. ISPs have used ways like traffic shaping to rate limit P2P applications (e.g. Packeteer), filter P2P traffic or reduce the priority of P2P traffic, and so on. However, this causes the debate of 'network neutrality'. As discussed in [12], in most cases, network neutrality is welfare superior to any feasible scheme for prioritizing or tiering content. Traffic discrimination may also raise legal and ethical issues. For instance, in 2008, the Federal Communications Commission took action to reprimand Comcast for the blocking of P2P uplink traffic [46]. To avoid filtering, ISPs may consider other technical approaches such as ISP-friendly P4P [28] and dynamic bandwidth allocation protocols [27]. However, encryption and dynamic port forwarding make detecting P2P traffic harder, which the more popular P2P systems are migrating to.

Without violating network neutrality, ISPs are left with few options to maintain profitability when facing the capacity pressure from increasing broadband access, especially from P2P applications. In the absence of effective technical solutions to solve the problem, economic responses are necessary. One possibility is to change the popular flat-rate pricing. In the following, we will explore a temporal-based pricing model that can be welfare enhancing for both ISPs and end-users.

4. A TEMPORAL-BASED PRICING MODEL

Rather than attempt to combat an ever-moving target, the alternative we consider is to address the fundamental economics of the situation. In particular, we propose a specific temporal-based pricing model that considers both 'all-you-can-eat' bandwidth selling at flat rates and usage-based charging based on time frame of use. Through our theoretical analysis, we show how a mutually beneficial model can be found that benefits both ISPs and end-users and is P2P friendly in the meantime.

4.1. Model assumptions

Throughout the analysis, we assume a fixed bottleneck capacity of the ISP and seek for a way the ISP can maximize profit without sacrificing QoS and at the same time allow the prosperity of P2P. Some other assumptions and parameters are described in the following:

- All users are potential P2P users. P2P and non-P2P applications have different access patterns. The non-P2P applications have an on/off pattern, whereas P2P applications act as servers that continuously serve content to others via uplink. Assume on average, at any point in time, users have the probability ξ to participate in P2P applications and the probability of ρ to participate in non-P2P actions. Because P2P applications tend to request an always-on service in order to grab an increasing amount of content at a large size, we have $0 \leq \rho < 1$ and $0 \leq \xi \leq 1$.
- Because users' actual bandwidth occupation cannot exceed their subscription, we assume on average that P2P applications occupy η share of subscribed bandwidth and non-P2P applications occupy θ share of subscribed bandwidth. Because P2P applications tend to download and upload large digital content, we have $0 < \theta < \eta \leq 1$.
- The threshold of the lowest bearable QoS is defined as a percentage of the bandwidth subscription. For example, if users can bear a 40% loss of speed, the threshold will be 60% of the bandwidth subscription. Let β be the average tolerance rate of users, $\beta = 0.6$ in the example.

4.2. Users' optimization problem

We consider one standard measure of social welfare, that is, consumer surplus, that measures the difference between users' valuation of the service and the price paid to access the service. In particular, we define the term 'user surplus' as a measure of welfare for user i that is the difference between user's utility and subscription rate, that is,

$$Surplus_i = U(\xi_i) - P(b_i) \tag{2}$$

where the utility function $U(\xi_i)$ measures the value user i receives from the Internet access (including P2P participation) that increases with the bandwidth consumption of the user, positively related to the P2P participation rate ξ. $P(b_i)$ is the user's total payment to the ISP for the services that increases with the subscribed bandwidth. Note that this general pricing formula is applicable to various types of pricing model because even under flat rates, the monthly rate increases with rising bandwidth subscription. Because $P(b_i)$ is set by the ISP, the user's key strategy is to choose the bandwidth subscription (b_i) and the P2P participation rate (ξ_i) that maximize user surplus.

4.3. ISPs' optimization problem

Being profit-driven, the ISP's objective is to maximize profit by choosing an optimal pricing scheme and setting corresponding optimal access rates. Because price is a key determinant of users' access behaviour, via optimal pricing strategy, the ISP aims at a manageable P2P participation rate ξ that prevents its service from failure and brings in the most possible revenue. As a baseline, we first derive the maximum profit of the ISP in the absence of P2P applications under the current flat-rate pricing model.

Because the down cost associated with the ISP is uncertain and can be noticeably large, one major objective of the ISP is to keep the QoS above the bearable threshold of users. To prevent traffic jams on the bottleneck capacity, there is an upper bound on the number of users the ISP can support without incurring down costs (or losing customers to competitors) at any fixed bottleneck capacity.

Suppose that the subscriptions to various bandwidth tiers by users are evenly distributed and $\bar{b} = (\sum_{i=1}^{N} b_i)/N$ is defined as the average downlink bandwidth subscription. To prevent the QoS from falling below the threshold, all feasible number of users the ISP can manage satisfies the following relationship:

$$\beta \cdot \rho \cdot N \cdot \theta \cdot \bar{b} \leq C \tag{3}$$

Thus, the maximum/equilibrium number of users the ISP can support at fixed resources is

$$N^* = \frac{C}{\beta \cdot \bar{b} \cdot \rho \cdot \theta} \tag{4}$$

The profitability of the ISP in the baseline case of a P2P-free scenario is

$$\pi^* = \sum_{i=1}^{N^*} P(b_i) - L(\overline{C}) \tag{5}$$

where $\overline{C}$ refers to the fixed capacity of the ISP (thus, the total cost of the ISP is fixed). Because the profit level corresponds to the maximum number of users the ISP can support without paying down costs, at flat rates, revenue has been maximized; hence, profit is also maximized for the ISP. This baseline case is the best the ISP can do given limited resources at flat rates.

We emphasize that the ISP's profit is maximized at N^* because it brings in the maximum amount of revenue at a fixed cost. Because pricing affects users by changing their incentives, any change in the pricing model forced by the threat of heavy P2P applications ought to minimize the impact on normal users while imposing financial disincentives on heavy P2P users; that is, by facing extra fees for their overconsumption of bandwidth, P2P users can no longer be completely subsidized by normal users. Their service access rates would be contingent on their actual consumption of bandwidth.

Of various pricing models, usage-based pricing is often discussed as opposed to flat rates. Usage-based pricing can take different formats, from strict charging by bytes to usage tiers. For any usage-based type of pricing to compete with flat rates, consumer resistance has to be dealt with as

consumers are protective of their broadband access. For example, in 2011, the Canadian Radio Television and Telecommunications Commission (CRTC) re-examined all usage-based policies in response to consumer outrage sparked by the January 2011 CRTC decision [47]. Carefully designed to avoid unnecessary impact on normal users, a temporal-based pricing model that is close to a combination of flat rates and usage-based pricing is suggested. The optimal solution to the pricing scheme leads to an optimal P2P frequency ξ^* that results in a win-win situation: at ξ^*, the ISP's profit is maximized, and P2P applications reach the maximum manageable level under the ISP's current bottleneck capacity.

4.4. Temporal-based pricing

We will lay out the basics of the pricing scheme through a simplified version of the ISPs' business model. The idea is that the ISP allows users to choose from tiered services and charges by time usage based on the bandwidth subscription for Internet access beyond a set cap. By finding an optimal rate to charge per unit time of use (e.g. $1 per day beyond the cap), the ISP can guarantee the subscribed bandwidth to all users and, in the mean time, maximize the profit and user surplus of both the ISP and end-users. The logic is as follows:

- Users purchase bandwidth from their ISPs and participate in P2P networks. The size of the bandwidth subscription and the frequency of P2P participation users choose are the results of their optimal decision-making.
- Following the preceding statement, the utility of users shall be maximized if users are guaranteed the subscribed bandwidth at their desired time of use.
- Payment to the ISP affects users' welfare. Specifically, a user's desired access request depends upon the price charged by the ISP.
- The ISP may limit a user's desired access with its pricing strategy.
- Charging by bandwidth per time of use may prevent the ISP's bottleneck bandwidth from being congested by P2P or other bandwidth-intensive applications.

Before moving on, we emphasize that a bandwidth cap can be imposed on the total data volume one may use in a month, for example, 60 GB/month. To minimize the possibility that the cap leads to surcharge on normal users, ISPs may adjust the cap, based on their best knowledge of the regular data consumption by normal users, so that the cap will apply only to heavy users. That way, normal users will still enjoy 'all-you-can-eat' access. For heavy users, they may avoid the surcharge by restricting P2P participation. The temporal-based pricing beyond the cap has the signalling effect on heavy users to limit their bandwidth occupancy.

In particular, the pricing mechanism works as such: ISPs set a monthly bandwidth cap. If the actual data consumption of a user is no higher than the cap, the user pays a flat monthly rate, just as in the current business model. If the monthly data usage exceeds the cap, the user will be charged additional fees on a temporal basis, for example, daily. It is not necessary for ISPs to choose daily time windows although daily surcharges appear a natural choice in temporal-based pricing. For instance, if the user reaches the bandwidth usage cap on the 25th day of the month, then for the remaining 5 days of the month, the user will pay daily extra charges for Internet access. The extra cost is temporal-based, rather than per byte or usage based. The simplest possible way to set the daily rate is $p(b_i) = P(b_i)/30$, the regular monthly flat rate divided by the number of days in a month.

4.5. A simple starting point

Suppose we start with homogenous users in the sense that they all purchase the same amount of bandwidth (equivalent to the ISP offering no tiered services), then $b_i = \bar{b}$, $\forall i = 1, 2, \ldots, N$, where N is the number of users.

Because the down cost for the ISP is uncertain and can be noticeably large if incurred (which may even cause the failure of the ISP), a key task of the ISP is to prevent its QoS from falling below the threshold. Given the number of users, the maximum/optimal P2P participation rate ξ^* is solved as

$$\beta \cdot \xi^* \cdot N \cdot \eta \cdot (1+\alpha) \cdot \bar{b} = C \tag{6}$$

where $\eta \cdot (1 + \alpha) \cdot \overline{b}$ is the average bandwidth requirement by P2P participation of both downlink and uplink bandwidths.

By solving (6), the maximum P2P participation rate feasible for the ISP is

$$\xi^*_{simple} = \frac{C}{\beta \cdot N \cdot \eta \cdot (1 + \alpha) \cdot \overline{b}} \tag{7}$$

Note that this simple starting point highlights some major factors determining the manageable P2P demand such as the ISP's bottleneck capacity (measured by C), users' QoS tolerance (measured by β) and bandwidth occupation by P2P (measured by η). The ISP would bear higher P2P penetration if it expands bandwidth capacity (an increase in C), if users can tolerate even worse QoS (a decrease in β) or if the average bandwidth requirement by P2P applications is reduced (a decrease in η).

4.6. Heterogenous users and user surplus maximization

In reality, ISPs sell various products, and the mix of users is heterogenous. Normally, tiers of bandwidth options are available for users to choose from. For generalization, we ignore the groups of tiers and assume an individually heterogenous bandwidth subscription.

We redefine $p(b_i)$ as an arbitrary unit time price set by the ISP that increases with the subscribed bandwidth. For illustration purpose, days are still used as the per unit time of pricing. At an appropriately chosen bandwidth cap, normal users are not subject to the additional daily fees. For heavy P2P users, by paying the additional daily rates for the remaining days after reaching the monthly cap, the total bandwidth consumption within a day is still unlimited up to the bandwidth subscription. Following marginal analysis, the consumer surplus of P2P users reaches the maximum when the marginal cost of an additional day of Internet access is equal to the marginal utility that extra day of P2P participation brings to the user, that is, $MU_i = p(b_i)$, where MU_i is the marginal utility and $p(b_i)$ is the marginal cost, that is, daily cost.

Figure 1 illustrates how user i's consumer surplus is maximized. Two curves are relevant in this figure: the marginal utility (MU) curve and the daily cost curve. The former shows diminishing marginal utility of beyond-cap Internet access; that is, each additional day of overconsumption has a lower marginal utility. The latter is horizontal, reflecting a constant daily surcharge. Because the horizontal axis is the length of time ranging from 0 day (no beyond-cap access request) to 30 days (a theoretical extreme), the vertical distance between the marginal utility curve and the daily cost curve is the user surplus on any day that surcharge is paid. In the range where the marginal utility curve is above the cost line, the user surplus is positive, and increasing bandwidth consumption by P2P participation will increase the total user surplus. In the range where the marginal utility curve is below the cost line, increasing the length of P2P use will make the user worse off. The optimal P2P participation is determined by the intersection of the two curves (denoted as d_i^*), which maximizes the total surplus of the user from P2P participation. In other words, at the daily rate of $p(b_i)$ for beyond-cap bandwidth consumption, the optimal P2P participation by user i (ξ_i^*) results in d_i^* days of surcharge.

Similar to the determination of P2P participation by user i, the likelihood of P2P participation for all users can be figured out at any given per-day surcharge set by the ISP. The optimal P2P participation probability chosen by individual users maximizes their total surplus.

4.7. Optimal price and profit maximization

By imposing additional temporal-based rates, ISPs can effectively avoid the jamming of bottleneck capacity by preventing P2P from consuming too much capacity. In the preceding analysis, we use a simple way of setting daily charges. Nevertheless, what are the general rules of price setting so that ISPs can reach the maximum profit level?

As discussed, users' P2P participation strategy depends upon the price charged by the ISP. In principle, all feasible P2P participation ξ_i satisfies
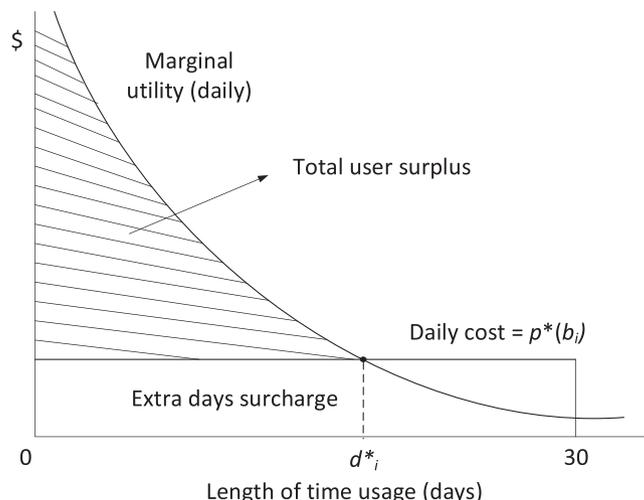
Figure 1. User optimal choice of bandwidth consumption under the temporal-based pricing model

$$\sum_{i=1}^{N}(\beta \cdot \xi_i \cdot \eta \cdot (1+\alpha) \cdot b_i) \leq C \tag{8}$$

where users are free to choose their individual bandwidth subscription (b_i) and, at any point in time, the actual bandwidth consumption of all users combined cannot exceed the bottleneck capacity of the ISP. In this set-up, heterogenous P2P users differ in their bandwidth subscription and how actively they are involved in P2P applications. They have the same QoS tolerance rate, and their P2P participation occupies the same percentage of bandwidth as these two are exogenous: the former is mainly psychologically determined, and the latter depends on P2P techniques.

The optimal solution requires a set of ξ_i^* so that

$$\sum_{i=1}^{N}\left(\beta \cdot \xi_i^* \cdot \eta \cdot (1+\alpha) \cdot b_i\right) = C \tag{9}$$

We refer to p^* as the optimal unit time price that would maximize the profit of the ISP. Similar to the determination of d_i^* in Figure 1 at any arbitrary price p, we can estimate the frequency of P2P access by a representative user i in response to p^*. Suppose at p^*, user i participates in P2P that costs d_i^* days of surcharge, and such a set of P2P participation (ξ^*) by all users solves equation (9), then the price p^* would be the optimal price.

At p^*, user i's daily payment to the ISP is $p^*(b_i)$. At such a price, the ISP's profit is improved compared with the pure flat rates because on one hand, the ISP still charges monthly flat rates on all its users, and in addition, heavy P2P users pay extra fees for beyond-cap data consumption on a daily basis; thus, the ISP receives more revenue. By financially discouraging heavy P2P usage of the bandwidth capacity, the ISP does not have to expand existing bottleneck capacity, so its costs would not increase. Users' actual P2P consumption and the corresponding participation time are the results of users' utility maximization problem.

Figure 2 illustrates the optimal byte consumption choice by a heavy user (the large inverse U-shaped curve) and a normal user (the small inverse U-shaped curve) under temporal-based pricing (and flat-rate pricing). User surplus at a given level of byte usage is the vertical distance between the user's total utility and the subscription rate, as in equation (2). Under flat rates, the largest distance between the inverse U-shaped total utility curve and the horizontal flat cost curve occurs at the satiation point, the turning point of the inverse U. That is, the heavy user chooses an optimal byte usage level of B_{flat}. Under temporal-based pricing, the cost curve becomes step shaped beyond the bandwidth cap. The
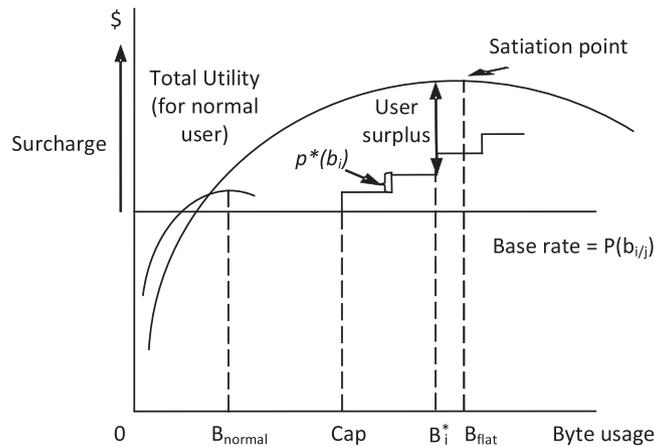
Figure 2. Optimal bandwidth consumption under temporal-based pricing

height of the steps is the optimal daily surcharge of $p^*(b_i)$ set by the ISP. The width of the steps is the average daily byte usage by the heavy user on the charged-per-time days. Surcharges on 4 days are plotted. In Figure 2, B_i^* is the optimal byte consumption that maximizes user surplus for the heavy user. Note that for the normal user, the optimal byte usage is of B_{normal}, that is, below the bandwidth cap regardless of the pricing scheme. In Figure 2, the relative curvature and position of the two total utility functions are for illustration purposes only.

The total revenue of the ISP is

$$R = \sum_j P(b_j) + \sum_i \left\{ P(b_i) + p^*(b_i) \times \frac{B_i^* - Cap}{c_i} \right\} \qquad (10)$$

where normal users (j's) pay only the base rate and heavy users (i's) pay the base rate plus surcharge if any. c_i is the average daily byte usage by the heavy user on charged-per-time days. $\frac{B_i^* - Cap}{c_i} \geq 0$ is the number of days with the surcharge, equivalent to d_i^* in Figure 1. At a fixed cost, ISP profitability increases under temporal-based pricing, compared with flat rates.

In summary, the temporal-based pricing model provides a win-win solution to the problem imposed by P2P applications on the scarce bottleneck capacity of ISPs. At the optimal unit time price, not only does the ISP receive more profit, but P2P also reaches the highest affordable flow given the ISP's bottleneck capacity level.

5. DISCUSSIONS

5.1. Flat rates and bundling

The Internet has historically treated all packets equally, and pricing has been predominantly through flat monthly rates depending only on the size of access links, not on usage. One main argument for flat-rate pricing is that flat rates reduce transaction costs of the users. The predictability of money flows is attractive for both ISPs and users.

In particular, flat rates for ISPs represent a form of bundling that offers access to a large number of web sites and services for a single price. It makes it convenient for users by providing a single bill and a single point of contact. For most reasonable distributions of demand, bundling is better for the service providers. However, bundling is not always more profitable than separate selling. Flat rates are very effective in stimulating usage, which helped the rapid growth of the Internet in early days. With the exponential increase in the data volume exchanged over the Internet, it makes sense for network operators and researchers to consider how to make efficient and effective use of Internet resources. Flat rates have generally been regarded as irrational and economically and socially undesirable. They are better for ISPs than metered rates only if the cost of carrying the additional traffic in the flat-rate

scenario is negligible. In the case that ISPs are forced to expand capacity to serve the increased demand for bandwidth, the marginal cost is not trivial. The extra cost of paying transit ISPs for larger bottleneck capacity may exceed the additional revenue from the increased number of subscribers attracted by simple flat rates. When ISPs start differentiating traffic or even block certain applications, flat rates are no longer beneficial for either party [40].

When marginal costs of serving increased demand for bandwidth are not trivial, as in the case of bottleneck congestion, price and quality differentiation can provide higher revenues and increase utilization efficiency of a network and thus in general increase social welfare, but neither price discrimination nor traffic discrimination is easy to implement in reality. Traffic discrimination violates the principle of network neutrality and faces possible legal issues that are not in the direct control of ISPs. Relatively, price differentiation can be unilaterally determined by ISPs because ISPs have the right to set and adjust their pricing models. Facing the increased pressure on bottleneck capacity and thus profitability, there are strong incentives for ISPs to engage in price differentiation.

5.2. Usage-based pricing

There are various types of usage-based pricing. Metered rates or charging by bytes as opposed to all-you-can-eat pricing model are often discussed and were once experimented by ISPs [14]. Usage-based billing may be attractive but can be hard to work out for user-preference and marketing reasons. The big challenge on any effort to break the all-you-can-eat model is the uncertainty of consumer responses. While ISPs may prefer fine-scale metering, customers fight for simplicity. Any attempt to change the flat-rate pricing will have to take into account such user preferences. As it has long been recognized, there exists strong public preference for flat-rate pricing. It is highly unlikely for ISPs to adopt strict usage-based solutions such as metered rates, especially if competing ISPs stay with the existing flat-rate pricing model.

Flat rates bring convenience to users at the cost of efficiency. Access pricing is subject to a continuing conflict between the need to optimize and people's reluctance to optimize. Any feasible changes in the pricing model cannot deviate much from the conventional flat rates. The temporal-based pricing model as a hybrid of flat rates and metered rates may be an applicable approach to benefit both ISPs, users and P2P applications.

5.3. Temporal-based pricing

One question naturally follows; that is, how effective is the temporal-based pricing? First of all, the temporal-based pricing model can work at least equally well in terms of affecting bandwidth consumption as metered rates because the two pricing schemes may lead to the same results if appropriate rates are chosen.

Figure 3 illustrates charging by bytes. The convex curve is the marginal utility of beyond-cap byte consumption, and the horizontal curve shows the per-byte price. The intersection of the two curves determines the optimal beyond-cap usage by the consumer, denoted as B^*. As in Figure 1, if B^* bytes require d^* days of access, then the corresponding per-byte price in the usage-based model and the unit time price in the temporal-based model would be equivalent as they generate the same amount of bandwidth consumption. For comparison, at flat rates, users will keep increasing their bandwidth consumption until their marginal utility reaches zero, so their bandwidth consumption choice is B_{flat} as shown in Figure 3.

Compared with strict charging by bytes, the temporal-based model has a major advantage in the sense that it accommodates rather than kills P2P by minimizing the cause for the churn effect. Under charging by bytes, end-users would get offline immediately after any downloading beyond the usage cap. Under the temporal-based pricing model, daily, weekly or other time window pricing can be used (although as an example, a daily window is used in the previous modelling analysis). Because users pay additional fees per unit time of usage, within the time frame, Internet access is still 'all-you-can-eat', favouring P2P users if wider window sizes are used. Users would be enticed to stay connected, thus reducing the churn rate of P2P. Alternatively, the current 'all-you-can-eat' pricing offers a 30-day window to users at a monthly rate. Charging by bytes narrows the window to none. Depending on the
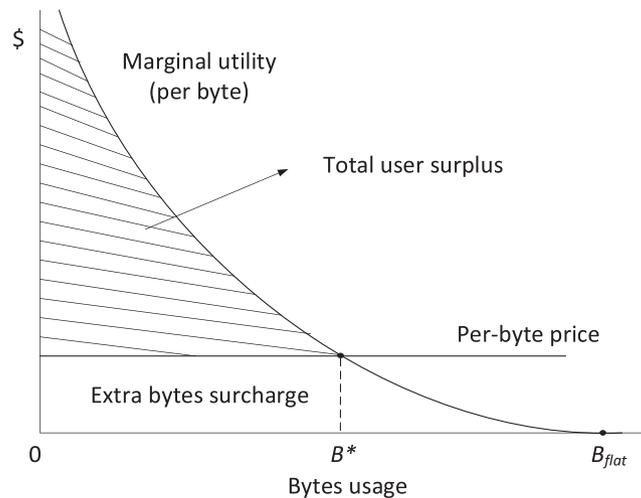
Figure 3. User bandwidth consumption under the per-byte charging model and flat-rate model

chosen time window, temporal-based pricing has the flexibility of adjusting the window size: the wider is the window, the more favourable it is to P2P.

It is much easier to market the temporal-based pricing to consumers than charging by bytes. The pricing model targets heavy users only. ISPs should be careful at setting the usage cap to minimize the possible influence on normal users. For example, ISPs can use their statistics of network traffic flow to set a cap that is above the actual monthly bandwidth consumption by normal users. Because the additional fees occur only after users hit the bandwidth cap, normal users would be indifferent under either the flat-rate pricing or the temporal-based pricing model. To avoid unnecessary negative responses from normal users, it is essential for ISPs to communicate the pricing idea to users. The Internet market is subject to the problem of asymmetric information. In terms of bandwidth consumption, users are less informed as they do not generally know their bandwidth consumption. ISPs can provide the information. One possible solution is for ISPs to do something similar to utility companies such as electricity companies that include monthly electricity usage history in the bill for the past 12 months or so. With the bandwidth consumption record in hand, normal users can easily see that their bandwidth consumption is (far) below the bandwidth cap so that they will be indifferent.

By charging first per volume and then per time, heavy users may be incentivized to concentrate their P2P traffic on the charged-per-time days, resulting in higher traffic towards the end of a month. To avoid the potential side effect, overlapping 30-day periods on individual basis can be used rather than using natural months to disperse traffic across time.

In summary, the temporal-based pricing model provides an economic rather than technical solution to the conflicts between ISPs' fixed resources, flat-rate pricing and the increased traffic demand from heavy users, especially heavy P2P users. It allows both ISPs and users to use the current software, protocols and hardware. The temporal-based pricing leads to a mutually beneficial solution that increases the profit of ISPs and users' utility surplus. By combining the advantages of flat rates and metered rates, it prevents the risks carried by the 'all-you-can-eat' pricing model and also accommodates P2P, rather than killing it as the simple usage-based pricing model does. By setting an appropriate bandwidth cap that is above the regular bandwidth consumption by normal users, the temporal-based pricing model leaves normal users unaffected and targets only heavy users. It provides incentives for heavy users to limit their bandwidth consumption, which are absent under flat rates. With transparent and sufficient information provided by ISPs to users, well-informed users will respond rationally to the adjustment in the pricing scheme.

Adopting the proposed pricing scheme by ISPs is less related to the heated debate on the network neutrality because no rate limiting, traffic policing or shaping on P2P applications is involved. The focus of the pricing scheme is to change users' economic incentives.
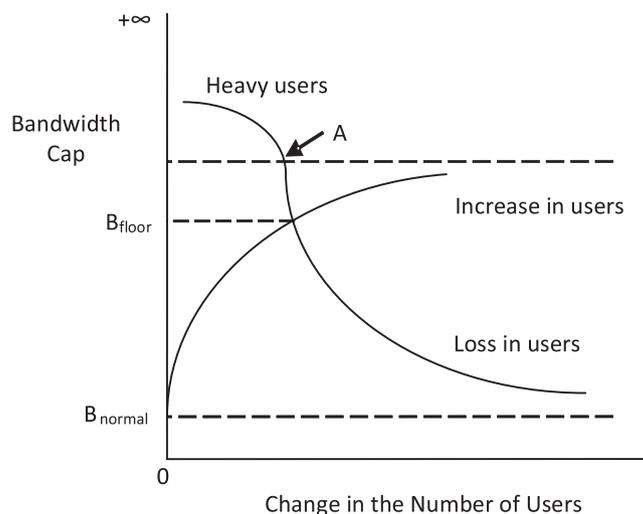
Figure 4. The impact of bandwidth cap on the changes in user body

Even in a competing environment where some ISPs adopt the proposed pricing scheme while others stay with the flat-rate model, it is unlikely for the pioneer ISPs who shift first to temporal-based pricing to lose customers other than those heavy users. Indeed, by losing heavy users, the ISPs can serve normal users with better QoS, which can even attract more normal users. Figure 4 illustrates the changes in the size and composition of an ISP's customers as it shifts from flat rates to temporal-based pricing, while its competitors stay with the conventional pricing scheme. The ISP starts with a bandwidth cap of infinity under flat rates. As it shifts away from flat rates (moving down along the y-axis), it starts losing customers to competitors. The inverse S-shaped curve shows the decrease in the number of users as the bandwidth cap falls. Heavy users dominate the loss in users initially as they are more bandwidth cap sensitive. Beyond the turning point A, it is the normal users that dominate. The other curve is the increase in the number of normal users choosing the ISP, attracted by its superior QoS. B_{normal} is the regular byte usage by normal users, which is generally much lower than any considered cap [15]. B_{floor} is the lower bound of the bandwidth cap that prevents the ISP from a net loss in the number of customers.

As heavy users migrate among ISPs, the ISPs staying with the conventional flat rates will face intensified bottleneck backbone pressure. Because the down cost may be very large, other ISPs will be induced to deviate also from flat rates. As more, and even all, ISPs adopt the temporal-based pricing scheme, the scheme is more likely to succeed.

5.4. Limitations

Like many theoretical analysis, our economic model is based on a few assumptions. One particular simplicity in our analysis is that we consider only the traffic that uses the transit link purchased by a local ISP from its upstream transit ISPs. With the technology of P4P, P2P traffic does not have to go through the transit link when the seeds are local. Technically, P4P [28] is a geographically aware P2P system where ISPs provide topology information to guide the P2P peering and sharing of content. A typical P2P model may have data sent across the Internet, outside of the local ISP, whereas P4P assigns priority to nearby peers and tries to download from the local nodes first. Therefore, P4P may save ISPs some costs if both peers are users of the same ISP because the traffic within the ISP does not cost bottleneck capacity. If P2P is no longer neutral but rather assigns a higher priority to local peers (P4P), then intuitively, as the number of P2P users increases, it will become easier for users to find content locally. With a significant number of P2P users within an ISP, higher P2P participation may mean more local seeds so that P2P can possibly help reduce traffic pressure on transit links. Thus, P4P may change the nature of the ongoing battle between P2P and ISPs. In future work,

we will explore the complex implications P2P may have on ISPs by taking into account P4P, P2P free riding and other related details of P2P characteristics and seek for optimal strategies of ISPs in various scenarios.

In addition, imposing monthly data allowance is not a completely new idea. While all-you-can-eat broadband access is predominant in home Internet services, it has been common for mobile phone operators to place caps on the monthly usage of cellular data plans. Some ISPs have also started placing monthly data caps. For instance, Comcast put its bandwidth cap of 250 GB in place in 2008. The decision was aimed at a small number of Internet users who Comcast felt were abusing their all-you-can-eat privileges by downloading a steady stream of high-definition movies from P2P networks. The consequences of exceeding the cap differ by ISP: some impose a surcharge per unit volume over the cap, some reduce the throughput of heavy users and so on. The surcharge discussed in this paper is temporal based. We note that the major contributions of this paper lie in its analysis of the temporal-based pricing mechanism beyond the bandwidth cap and its attempt to explore mutually beneficial pricing mechanisms in the face of the increasing tension between ISPs and Internet users introduced by emerging applications and technologies and the tremendous data exchanged. The temporal-based pricing model has nice features such as marketability and flexibility, which allows ISPs to choose an appropriate time window depending on their own conditions and priorities. Sliding windows can also be readily accommodated to disperse data flows. Because the difficult situation in pricing Internet services results from various conflicting factors, we are positive that in the future, there will be a mix of pricing models that varies depending on technologies and local conditions.

6. CONCLUSION

As we move into the 'big data' era, more and more data are being shared and exchanged over the Internet. The explosion of bandwidth consumption by end-users and the different traffic patterns of new emerging applications put a significant load on ISP bottleneck link capacity and have been forcing ISPs to reconsider their current business model of overselling the all-you-can-eat broadband access at monthly flat rates. These challenges are essential to the business sustainability of ISPs and are in turn critical to the prosperity of data networks, especially P2P networks because P2P applications by nature serve as both client and server and tend to generate more data in both directions. In this paper, we study a temporal-based pricing model in seeking for a mutually beneficial solution to, on one hand, release the rising tension between ISPs and heavy users in general by changing financial incentives of both parties and, on the other hand, accommodate the prosperity of P2P applications rather than killing them. Through characterizing the impact of the increasing bandwidth pressure on the ISP business model and performing an economic analysis, we demonstrated how a temporal-based pricing scheme can benefit both ISPs and Internet users, in particular P2P networks, in a win-win situation.

REFERENCES

1. Shakkottai S, Srikant R. Economics of network pricing with multiple ISPs. *IEEE/ACM Transactions on Networking* 2006; **14**(6): 1233–1245.
2. Parameswaran M, Susarla A, Whinston AB. P2P networking: an information sharing alternative. *Computer* 2001; **34**(7): 31–38.
3. Krishnan R, Smith MD, Telang R. The economics of peer-to-peer networks. *Journal of Information Technology Theory and Application* 2003; **5**(3): 31–44.
4. Saroiu S, Gummadi PK, Gribble SD. A measurement study of peer-to-peer file sharing systems. In *Proceedings of Multimedia Computing and Networking 2002,* San Jose CA, 2002.
5. Sandvine. Global internet phenomena report, 2013. 2H.
6. Rodriguez P, Tan S-M, Gkantsidis C. On the feasibility of commercial, legal P2P content distribution. *ACM SIGCOMM Computer Communication Review* 2006; **36**(1): 75–78.
7. Zheng X, Cho C, Xia Y. Optimal peer-to-peer technique for massive content distribution. In *Proceedings of the 27th Conference on Computer Communications,* Phoenix, AZ, 2008; 151–155.
8. Neglia G, Reina G, Zhang H, Towsley D, Venkataramani A, Danaher J. Availability in BitTorrent systems. In *Proceedings of the 26th IEEE Conference on Computer Communications,* Anchorage, AK, 2007; 2216–2224.

9. Labovitz C. The rise and fall of P2P. DeepField Networks, 2012. Available: http://www.deepfield.com/2012/04/the-rise-and-fall-and-rise-of-p2p/ [10 April 2012].

10. Jackson M. EU telecoms regulators find ISP blocking of P2P and VoIP is common. ISPreview, 2012. Available: http://www.ispreview.co.uk/index.php/2012/03/eu-telecoms-regulators-find-isp-blocking-of-p2p-and-voip-is-common.html [13 March 2012].

11. Beverly R, Bauer S, Berger A. The Internet's not a big truck: toward quantifying network neutrality. In *Proceedings of the 8th Passive and Active Measurement Conference,* Louvain-la-neuve, Belgium, 2007; 135–144.

12. Economides N, Hermalin BE. The economics of network neutrality. *RAND Journal of Economics* 2012; **43**(4): 602–629.

13. Economides N. The economics of the Internet backbone. In *Chapter 9 of Handbook of Telecommunications Economics*, Majumdar SK, Vogelsang I, Cave ME (eds.), Vol. 2 Elsevier (North Holland Publishing Co.): Amsterdam, Netherlands, 2006, pp. 373–412.

14. McElligott T. Usage-based billing back on agenda. Billing & OSS World, 2008. Available: http://www.billingworld.com/articles/2008/02/usage-based-billing-back-on-agenda.aspx [20 February 2008].

15. Lyons D. Internet pricing: The next policy frontier. All Things D, 2013. Available: http://allthingsd.com/20130401/internet-pricing-the-next-policy-frontier/ [1 April 2013].

16. Dian SM, Ramadoss B. Trading of grade based incentives to avoid free riding and starvation in P2P network. *Lecture Notes in Computer Science—Distributed Computing and Internet Technology* 2013; **7753**: 347–360.

17. Golle P, Leyton-Brown K, Mironov I. Incentives for sharing in peer-to-peer networks. In *Proceedings of the 3rd ACM Conference on Electronic Commerce,* Tampa, FL, 2001; 264–267.

18. Figueiredo D, Shapiro J, Towsley D. Incentives to promote availability in peer-to-peer anonymity systems. In *Proceedings of the 13th IEEE International Conference on Network Protocols,* Boston, MA, 2005; 110–121.

19. Jun S, Ahamad M. Incentives in BitTorrent induce free riding. In *Proceedings of the ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems,* Philadelphia, PA, 2005; 116–121.

20. Li M, Yu J, Wu J. Free-riding on BitTorrent-like peer-to-peer file sharing systems: modeling analysis and improvement. *IEEE Transactions on Parallel and Distributed Systems* 2008; **19**(7): 954–966.

21. Zhang H, Neglia G, Towsley D, Presti GL. On unstructured file sharing networks. In *Proceedings of the 26th IEEE Conference on Computer Communications,* Anchorage, AK, 2007; 2189–2197.

22. Habib A, Chuang J. Incentive mechanism for peer-to-peer media streaming. In *Proceedings of the 12th IEEE International Workshop on Quality of Service,* Montreal, Canada, 2004; 171–180.

23. Feldman M, Papadimitriou C, Chuang J, Stoica I. Free-riding and whitewashing in peer-to-peer systems. *IEEE Journal on Selected Areas in Communications* 2006; **24**(5): 1010–1019.

24. Ge Z, Figueiredo DR, Jaiswal S, Kurose J, Towsley D. Modeling peer-peer file sharing systems. In *Proceedings of the Twenty-Second Annual Joint Conference of the IEEE Computer and Communications*, Vol. 3: San Francisco, CA, 2003; 2188–2198.

25. Yan Y, El-Atawy A, Al-Shaer E. Ranking-based optimal resource allocation in peer-to-peer networks. In *Proceedings of the 26th IEEE Conference on Computer Communications,* Anchorage, AK, 2007; 1100–1108.

26. Zhang Z, Chen S, Yoon M. March: a distributed incentive scheme for peer-to-peer networks. In *Proceedings of the 26th IEEE Conference on Computer Communications,* Anchorage, AK, 2007; 1091–1099.

27. Xie H, Yang YR, SilberSchatz A. Towards an ISP-compliant, peer-friendly design for peer-to-peer networks. In *Proceedings of the 7th International IFIP-TC6 Networking Conference on AdHoc and Sensor Networks, Wireless Networks, Next Generation Internet,* Singapore, 2008; 375–384.

28. Xie H, Yang YR, Krishnamurthy A, Liu Y, Silberschatz A. P4P: provider portal for applications. In *Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication,* Seattle, WA, 2008; 351–362.

29. Guo L, Chen S, Xiao Z, Tan E, Ding X, Zhang X. Measurements, analysis and modeling of BitTorrent-like systems. In *Proceedings of the USENIX Internet Measurement Conferences,* Berkeley, CA, 2005; 35–48.

30. Kamvar SD, Schlosser MT, Garcia-Molina H. The EigenTrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th International Conference on World Wide Web,* Budapest, Hungary, 2003; 640–651.

31. Cox LP, Noble BD. Samsara: honor among thieves in peer-to-peer storage. *ACM Operating Systems Review* 2003; **37**(5): 120–132.

32. Gurbani VK, Hilt V, Rimac I, Tomsu M. A survey of research on the application-layer traffic optimization problem and the need for layer cooperation. *Communications Magazine* 2009; **47**(8): 107–112.

33. Seedorf J, Kiesel S, Stiemerling M. Traffic localization for P2P-applications: the ALTO approach. In *Proceedings of the IEEE Ninth International Conference on Peer-to-Peer Computing,* Seattle, WA, 2009; 171–177.

34. Seibert J, Torres R, Mellia M, Munafo MM, Nita-Rotaru C, Rao SG. The Internet-wide impact of P2P traffic localization on ISP profitability. *IEEE/ACM Transactions on Networking* 2012; **20**(6): 1910–1923.

35. Wang JH, Chiu DM, Lui JCS. Modeling the peering and routing tussle between ISPs and P2P applications. In *Proceedings of the 14th IEEE International Workshop on Quality of Service,* New Haven, Connecticut, USA, 2006; 51–59.

36. Shin SJ, Weiss MBH. Simulation analysis of QoS enabled Internet pricing strategies: flat rate vs. two-part tariff. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences,* Hawaii, USA, 2003; 1-10.

37. Dhamdhere A, Dovrolis C. Can ISPs be profitable without violating "network neutrality"?. In *Proceedings of the 3rd International Workshop on Economics of Networked Systems,* Seattle, WA, USA, 2008; 13–18.

38. Wu Y, Hande P, Kim H, Chiang M, Tsang DHK. QoS-revenue tradeoff with time-constrained ISP pricing. In *Proceedings of the 18th International Workshop on Quality of Service,* Beijing, China, 2010; 1–9.

39. Tran H-H, Tuffin B. Inter-domain pricing: challenges and possible approaches. *International Journal of Network Management* 2010; **21**(3): 223–246.
40. Odlyzko A. Internet pricing and the history of communications. *Computer Networks* 2001; **36**(5-6): 493–517.
41. Geng X, Stinchcombe MB, Whinston AB. Bundling information goods of decreasing value. *Management Science* 2005; **51**(4): 662–667.
42. Nabipay P, Odlyzko A, Zhang Z-L. Flat versus metered rates, bundling, and "bandwidth hogs". In *Proceedings of NetEcon 11: 6th Workshop on the Economics of Networks, Systems, and Computation,* San Jose, California, USA, 2011.
43. Ha S, Joe-Wong C, Sen S, Chiang M. Pricing by timing: innovating broadband data plans. In *Proceedings of the SPIE OPTO Broadband Access Communication Technologies VI Conference*, Paper 8282-12, San Francisco, California, USA, 2012.
44. McKnight LW, Leida B. Internet telephony: costs, pricing, and policy. *Telecommunications Policy* 1998; **22**(7): 555–569.
45. Sharma A, Silberman N, Subramanian L, Economides N. On the rise and fall of ISPs. In *Proceedings of Netecon 09: Workshop on the Economics of Networks, Systems, and Computation,* Stanford, California, USA, 2009.
46. Johnson F. FCC votes 3–2 to reprimand Comcast for slowing Internet. CNN Money, 2008. Available: http://money.cnn.com/news/newsfeeds/articles/djf500/200808011213DOWJONESDJONLINE000741_FORTUNE5.htm.
47. Buckley S. Canada's CRTC to hand down decision on usage-based billing this week. FierceTelecom, 2011. Available: http://www.fiercetelecom.com/story/canadas-crtc-hand-down-decision-usage-based-billing-week/2011-11-14.

AUTHORS' BIOGRAPHIES

Zhen Li is an Associate Professor of Economics in the Department of Economics and Management at Albion College. She holds a master's degree and PhD in Economics from Princeton University. She graduated with her bachelor's degree in International Economics from Peking University. Dr Li's recent research interests include inter-disciplinary research study in economics and game theory of computer networks and information security.

Qi Liao is an Assistant Professor in the Department of Computer Science at Central Michigan University. He received his MS and PhD in Computer Science and Engineering from the University of Notre Dame. He graduated with a BS and Departmental Distinction in Computer Science from Hartwick College with a minor concentration in Mathematics. His research interests include computer and network security, network management, data analysis and visualization, and economics of networking and cybersecurity. Dr Liao received awards including USENIX best paper award, second place in National Security Innovation Competition at Colorado Springs, University of Notre Dame Center for Research Computing Award for Computational Sciences and Visualization, and IEEE Visual Analytics Science and Technology Challenge Award. Dr Liao has served as an international conference co-chair, in technical program committees, journal editorial board and numerous peer reviews. He is a member of Kappa Mu Epsilon, Upsilon Pi Epsilon, and Tau Beta Pi. Dr Liao is an ASEE Faculty Fellow of US Air Force Summer Faculty Fellowship Program.