The Case for Microcontracts for Internet Connectivity

Vytautas Valancius*, Cristian Lumezanu*, Ramesh Johari†, Vijay Vazirani*
*Georgia Tech †Stanford University

ABSTRACT

This paper introduces *microcontracts*, which are contracts for "slices" of the Internet connectivity along dimensions such as time, destination, volume, and application type. Microcontracts are motivated by the observation that Internet service providers carry traffic for different classes of customers that use the ISP's resources in a variety of different ways and, hence, impose different costs on the ISPs. For example, customers have little incentive to move less important traffic from a peak time interval unless their contract reflects the ISP's costs in that time interval. To address this inefficiency, microcontracts divide connectivity into fine-grained units so that prices more directly reflect the costs that the ISP bears for delivering the connectivity at that time. We explore the feasibility of applying microcontracts in realistic Internet service provider settings by characterizing the traffic patterns from a transit network along two specific dimensions: time-of-day and distance travelled. We argue that microcontracts are both feasible and advantageous to both buyers and sellers of Internet connectivity. We develop a model to help ISPs derive customer demand functions from observed traffic patterns; using this model, we show that making contracts for Internet connectivity more fine-grained can improve the aggregate gain of an ISP and its customers.

1. Introduction

Internet connectivity depends traffic exchange between cooperating Internet Service Providers (ISPs) that establish contracts with each other. This cooperation led to the growth of the Internet accompanied by an increasing diversity of users, applications, access technologies, and networks. Unfortunately, the flexibility of Internet contracts has not kept pace with this increasing diversity: contracts for transit between ISPs mostly conform to "one size fits all", bulk connectivity or some variant (*e.g.*, peering, paid peering, or customer-provider contracts). A major shortcoming of this inflexibility is that the prices that users pay for connectivity do not directly reflect either the value that a customer has for the connectivity or the cost incurred by the ISP for carrying that customer's traffic.

In an attempt to address these shortcomings, we explore how to structure contracts to more directly reflect the costs that an ISP incurs for carrying a customer's traffic flows. Specifically, we introduce *microcontracts*, which are "slices" of Internet connectivity along different usage dimensions, such as time, destination, volume, or application. Unlike current "one size fits all" routing contracts, which trade connectivity in bulk, microcontracts offer finer granularity in traffic exchange. For example, an ISP could offer one microcontract for traffic sent during peak hours of the day and another for off-peak times. Such contracts enable the exchange of more fine-grained goods, ultimately making resource allocation more efficient and increasing the utility for customers and providers.

Our ultimate goal is to develop a framework to facilitate the implementation of microcontracts. Given traffic demands from customer networks and the associated infrastructure cost, an ISP could ultimately answer the following questions: (1) At what *granularities* and along what *dimensions* should it sell connectivity to its customers; (2) At what *price* should it offer these contracts to each customer? This paper does not completely answer these questions; we take a first step towards our goal, by building a connectivity market model and, backed by the data from a transit ISP, exploring the potential benefits of microcontracts for both buyers and sellers.

Some network providers already implement a rudimentary form of microcontracts called *tiered pricing*, where customers may pay different amounts depending on their network usage patterns. In 2008, Latin American cellular providers began to roll out differentiated pricing for "on-net" traffic (*i.e.*, traffic in their own networks) [1]. Recently, AT&T Wireless started charging different prices for cellular data plans, based on total traffic volume [9]. Even for transit connectivity, some ISPs offer differentiated pricing based on where their customers send the majority of their traffic [2]. Researchers have also studied the benefits of tiered pricing. Kesidis *et al.* [6] and Shakkottai *et al.* [8] studied the benefits of volume-based pricing. Jiang *et al.* [5] and Hande *et al.* [4] applied a tiered pricing model based on time.

In some sense, microcontracts generalize tiered pricing: with microcontracts, network service providers can slice contracts according to arbitrary granularities along a number of different dimensions where user behavior (and hence, the cost of a user) may vary. In particular, user behavior may vary according to time, destination, volume, and application type, as follows:

- *Time.* Some customers may use network predominantly during the day, while others may use it more at night.

- *Destination.* Some customers may be mainly interested in service to a small set of geographically close destinations, while others may primarily use international links and/or backbone.
- *Volume.* Some customers may send very little traffic, whereas other customers may send much larger volumes.
- *Application type.* Some customers may use applications that are far more sensitive to delay (*e.g.*, voice or video) than other customers, who may depend less on low delay or jitter.

Despite this variation, network providers do not commonly use tiered pricing, perhaps because it is difficult to determine how to set the appropriate granularities for contracts and the prices for each granularity. This paper takes a first step towards tackling this challenge.

This paper presents three contributions. First, we develop a general Internet connectivity market model to show that microcontracts can improve utility for both buyers and sellers of Internet connectivity. Second, we derive a "business case" for constructing microcontracts along the time and destination dimensions; this analysis mines network traffic from a large European Internet service provider network; we find that different adjacent customer networks have varying traffic patterns that use different network resources at different times. This non-uniformity reveals an opportunity for an ISP to structure contracts that more directly match each customer's demand. Finally, we show, using real network traffic data, how ISPs and their customers benefit from time and destination microcontracts, and that most aggregate benefit is achieved with few pricing intervals.

2. Modeling Microcontracts

In this section we present a case for microcontracts. We explore the following questions:

- Is an ISP better off if it sells microcontracts?
- Are the customers of that ISP better off?
- Under what conditions they are better off?
- How fine-grained should microcontracts be?

One of the main challenges in answering these questions is that current traffic demands only tell us about traffic volumes under the current contracts. First and foremost, we cannot tell what will happen to consumer demand if an ISP changes these prices. In other words, we do not know the consumer demand function. We also do not know how to estimate the ISP profit. Our model tackles these two challenges. We construct a model that captures how ISP profit and consumer demand vary with the price of connectivity in each slice (this section). We then evaluate microcontracts using our model and a traffic demand matrix from a large European ISP (Section 3). We show that using microcontracts that divide demand into smaller intervals by time or destination can increase profit for both ISPs and their customers.

2.1 Market Model

We first introduce a simple, general market model for Internet connectivity. Our model has two types of agents:

- **ISPs** that maximize *profit* by choosing the granularity of the Internet connectivity in various dimensions, and setting the prices in those dimensions. The ISP's profit is computed by subtracting costs from revenue derived from sold bandwidth.
- **Customers**, such as other ISPs, companies, or home users that purchase bandwidth. The users respond to price changes by reducing or increasing their traffic according to their *demand function*. The net payoff to consumers (utility of consumption minus payments) is called *consumer surplus*.

In addition to ISP profit and consumer surplus, we are interested in measuring *social welfare*: this is aggregate gain of all market agents.

We refer to the granularity of a microcontract in a particular dimension as a *pricing slice*. Consider a simple example, where an ISP structures its contracts with two pricing slices in the time dimension: one for day and one for night. The ISP can then set a higher price during the day when usage is higher, in an attempt to extract more profit and/or encourage users to use resources at night. We aim to determine whether such a strategy would leave ISPs and/or customers better off. Since market interaction is not a zero-sum game, in some cases all agents will benefit. In other cases, some agents will be made worse off, but as long as social welfare increases, it is in principle possible to design a payment mechanism that provides incentives for both ISPs and customers to deploy microcontracts.

2.2 Estimating the Effects of Microcontracts

We aim to model the effects of microcontracts on both consumer surplus and ISP profit, to determine whether users and ISPs are better off, respectively, as a result of microcontracts. Due to space constraints, in this section we only provide key definitions and the main assumptions of our methodological approach.

2.2.1 *Consumer Demand and Surplus*

To understand how the granularity and prices of microcontracts affect customers, we first model consumer surplus; to do so, we must first derive a function that represents how demand in a slice would change with the price for that slice. We derive a functional form for consumer demand by assuming that consumers are represented using an α*-fair utility* model, as in other models for demand in communication networks [7]. The demand function for this utility takes the following form; we omit the details due to space constraints:

$$Q_i(p_i) = a_i p_i^{\epsilon}, \quad (1)$$

where p_i is the price set by an ISP for slice i, $Q_i(p_i)$ is the quantity users buy in slice i given p_i, ϵ is the elasticity, and a_i is a demand coefficient. Elasticity determines how consumers respond to price changes and allows us to model the behavior of different classes of users; higher elasticity means that a user is more sensitive to changes in price.

Ultimately, consumer behavior depends on the elasticity, ϵ, and the demand coefficient, a_i, for each slice. Unfortunately, we do not know *either* of these parameters. To find the *unknown demand coefficients*, we fit the demand function to our traffic data, making two key assumptions. First, we assume that the ISP is currently setting a single price for all slices. Second, we assume that the ISP is currently maximizing revenue according to this constraint and the demand function. This second assumption allows us identify the optimal price, given the observed demand; we then use this optimal price and the observed demand to solve for the demand coefficients. (Given the optimal price $p*$ and the measured quantity q_i for slice i, the demand coefficient is simply $a_i = q_i / p*^{\epsilon}$.) The second problem is that *elasticities of users are unknown*. To solve this problem, we run experiments across a range of elasticities and observe how consumer demand varies in each case. Note that although we estimate different demand coefficients in each slice, we assume elasticity is the same in different slices.

2.2.2 *ISP Profit*

To model how ISPs will set prices for microcontracts, we must develop a model for ISP profit. ISP profit equals the revenue that it receives from selling connectivity minus its costs.

Revenue. An ISP's revenue in slice i is simply the price that the ISP sets for that slice, multiplied by the quantity that it sells at that price (as determined from the consumer demand function): $p_i Q_i(p_i)$.

Cost. An ISP's cost is more challenging to define. First, cost has several different aspects, each of which need to be modeled: an ISP's cost usually consists of operational expenses (OPEX), which depend on usage, such as leasing long distance capacity; and capital expenses (CAPEX) that include fixed costs such as infrastructure. Second, capacity use affects ISP costs in different ways, depending on the dimension. For example, with regard to time, shifting traffic to off-peak hours can increase revenue without increasing cost, since the traffic is carried on infrastructure that has been provisioned for peak usage. On the other hand, shifting traffic to different parts of the network infrastructure can affect costs significantly, since different network links and circuits incur different costs.

Let $\boldsymbol{q} = (q_1, \ldots, q_K)$ denote the vector of realized demands across slices. We let $C(\boldsymbol{q})$ denote the cost to the ISP for serving the traffic vector $\boldsymbol{q}$. In general, the cost is the total network cost to serve this traffic. For example, if $\boldsymbol{q}$ measures demands over time slices, then $C(\boldsymbol{q})$ is the total cost to serve the *peak* traffic, i.e., $C(\boldsymbol{q}) = c \max_i q_i$. Note that in the latter case, no additional cost is incurred if the peak provisioning is not exceeded; thus in particular, if the ISP is aware of a peak constraint in advance, then it can simply provision for it and choose prices to maximize revenue. On the other hand, if $\boldsymbol{q}$ measures demands over distance slices, then $C(\boldsymbol{q})$ might be modeled as a distance-weighted sum of costs across slices, e.g., $C(\boldsymbol{q}) = \sum_i c_i q_i$, where c_i increases with distance. Note that these are simple abstractions of the same underlying cost structure: provisioning sufficient network resources to meet demand.

The ISP's profit can thus be written:

$$\Pi(\boldsymbol{p}) = \sum_i p_i Q_i(p_i) - C(Q_1(p_1), \ldots, Q_K(p_K)), \quad (2)$$

where p_i is the price in slice i, and $Q_i(p_i)$ is the demand in slice i. Given this expression for ISP profit, we can solve for the optimal price necessary to compute the demand function coefficients for different slices.

3. Microcontracts in Practice

Summary. We estimate the benefits of microcontracts using real traffic data from a large European ISP. We present results for two dimensions: time and destination. For each dimension, we first analyze the usage patterns in potential pricing intervals, and show that they vary significantly across the intervals. This variation encourages the use of microcontracts: by applying different prices for each interval, an ISP can price connectivity more directly according to the costs that it incurs for carrying traffic for that interval. In some cases, it may also encourage its customers to shift traffic to intervals with less traffic, thereby reducing operational cost and increasing profit. We then describe general cases where both ISPs and their customers can benefit from microcontracts in each dimension.

Data. We use the traffic demand data from a UK transit ISP. The ISP transits approximately 1.4 petabytes of data a day, with peak at approximately 200 Gbit/s and off-peak at 50 Gbit/s. In this ISP, we identify 445 large customers who transport their traffic between seven major PoPs. We make no assumptions on the type of traffic that is exchanged. Our goal is to show that selling connectivity in slices based on time or destination is practical and can benefit both the sellers and the buyers.

3.1 Time Microcontracts

We explore the potential benefits of microcontracts in the time dimension. We ask whether both ISPs and customers can benefit if ISPs sell contracts for specific inter-

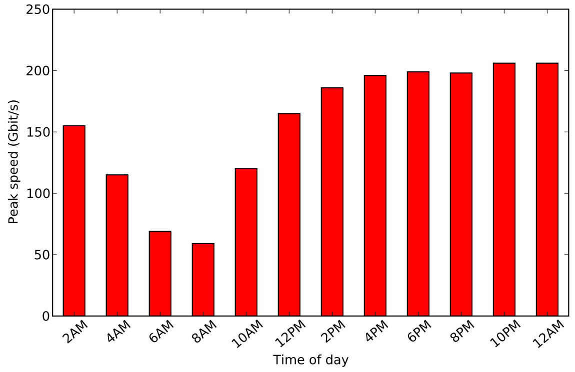

Figure 1: Maximum aggregate speed observed in 2-hour interval.

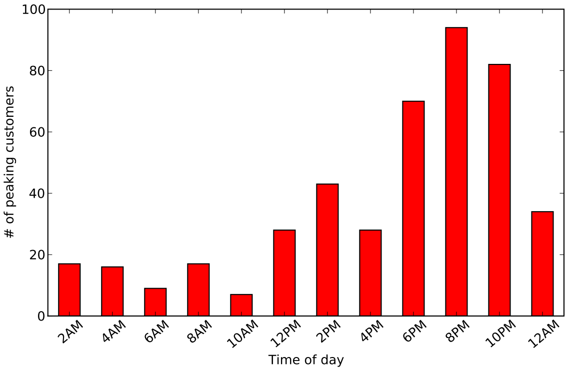

Figure 2: Number of users peaking in 2-hour interval.

vals of time, rather than in bulk. We begin with some observations about temporal patterns in network traffic. For customer traffic that is elastic, we show that time-based microcontracts can help ISPs reduce their peak load by encouraging customers to shift some of their traffic to time intervals that are less busy. We also observe that different customers send traffic at different times of day and, hence, that microcontracts can improve consumer surplus by more directly pricing the connectivity for these different customers according to its true cost.

3.1.1 *Observations*

Time-based microcontracts allow ISPs to reduce their peak load by encouraging more elastic customers to move traffic to periods with less load. Shifting traffic from these customers can directly reduce ISP's costs. Most of the traffic on the Internet exhibits a diurnal pattern. In addition, according to the Arbor Networks Atlas Observatory study [3], a large fraction of Internet traffic is non-interactive (*e.g.*, P2P traffic) or could be cached closer to customers (*e.g.*, VoD traffic), which makes it easier to time-shift it. Prices that more directly reflect the cost of carrying traffic during each time interval could encourage customers to shift some of this traffic to night and early morning time-slots, thus reducing load at peak times and, hence, the ISP's overall costs.

Figure 1 shows the rates at which ISP's customers download data. It plots the highest aggregate download rates in in a two-hour interval; the peak to off-peak ratio is about 4:1. Pricing traffic at higher rates during peak time intervals could encourage some of this traffic volume to shift to other less-busy time intervals, reducing peak utilization and using resources more efficiently during other times. Figure 2 shows the number of customers that peak at each one-hour interval; most peak between 3 p.m. and 11 p.m. Although our evaluation does not directly model this behavior, this observation implies that an ISP could offer different contracts to different users, depending on when they peak. Using these observations as a starting point, we quantify how time microcontracts can benefit ISPs and customers.

3.1.2 *When do time microcontracts help?*

If customer traffic demand were unbounded, an ISP could reduce prices to maximize utilization in all time intervals; this is unrealistic. Instead, we treat customers as having a finite set of tasks to accomplish (*e.g.*, uploading files, watching movies) that could be spread across different time intervals, with different preferences. To model this, we hold the *total* traffic volume constant ($C = \sum_{i=1}^{n} q_i$). In the context of the model, this implies that when an ISP increases the price for traffic in one time interval, it must reduce the price in another to maintain the same total traffic volume for that customer.

We can show that, in the context of our model, keeping the total traffic volume constant means that the original pricing already maximizes the ISP's profit (we omit this proof due to space constraints). Thus, we evaluate the effects of microcontracts in terms of an operational goal, such as reducing peak utilization or reducing traffic variance over the intervals. We then quantify how an social welfare changes as the ISP increases the number of time slices. In this paper, we focus on peak reduction.

Given a peak traffic load, K, over all time intervals, the ISP can attempt to reduce this peak load by some fraction $\beta < 1$ by selling microcontracts over different time intervals and pricing each interval to encourage customer demand during peaks to shift to different intervals. An ISP can determine the optimal pricing for each time interval by solving the following optimization:

$$\begin{aligned} \text{maximize} &: \sum_{i=1}^{n} p_i Q_i(p_i) \\ \text{s.t.} &: \sum_{i=1}^{n} Q_i(p_i) = C \\ \text{s.t.} &: Q_i(p_i) \leq \beta K, \forall i \in 1 \cdots n \\ \text{w.r.t.} &: p_i, \forall i \in 1 \cdots n \end{aligned} \tag{3}$$

Larger values of β reduce ISP profit and social welfare.

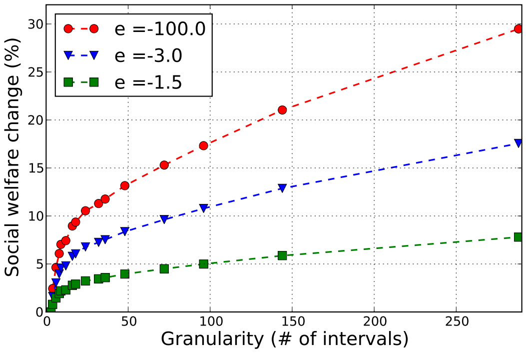

Figure 3: Social welfare increase as an ISP is increasing pricing granularity — both ISP profit and customer surplus is growing and contributing to it. ISP reduced peak utilization by 30% and generated higher social welfare when more pricing intervals were used. The plot shows results for different elasticities.

The relevant question in the context of microcontracts, however, is by how much this profit is affected when ISP increases the number of pricing intervals n. For example, using three intervals in the day, an ISP can set three different prices for intervals 12 a.m.–8 a.m., 8 a.m.–4 p.m., and 4 p.m.–12 a.m..

Intuitively, selling contracts on more time slices should allow the ISP to extract more profit for the same peak reduction. To quantify this gain, we solve above maximization using a simple heuristic: First, we assume that the ISP can sell contracts for n equal length intervals. We then find intervals that exceed βK and set the price point in these intervals so that resulting demand is reduced to βK. Then, we take the volume we just "eliminated" and apply this volume to underutilized intervals, starting with the most utilized interval below βK level, by reducing the price in those intervals according to the demand function. We then explore how increasing n (*i.e.*, selling more fine-grained microcontracts) affects social welfare.

Figure 3 shows that *increasing the number of time intervals increases social welfare*. In fact, both ISP profit and customer surplus contribute to this increase. As expected, the more elastic customer traffic is, the higher the social welfare (since the ISP loses less revenue trying to provide incentive for users to move traffic to different times). Additionally, most of the additional profit that an ISP gains by applying microcontracts can be extracted with a relatively small number of pricing granularities. For example, using 12 two-hour time intervals yields an 8% increase in profit, which is more than half of the additional profit that an ISP would achieve with 48 30-minute time intervals. This result is promising for implementing time-based microcontracts in practice: it suggests that there is a "sweet spot" whereby an ISP can increase its profits significantly without incurring the overhead of introducing arbitrarily fine-grained microcontracts.

3.2 Destination Microcontracts

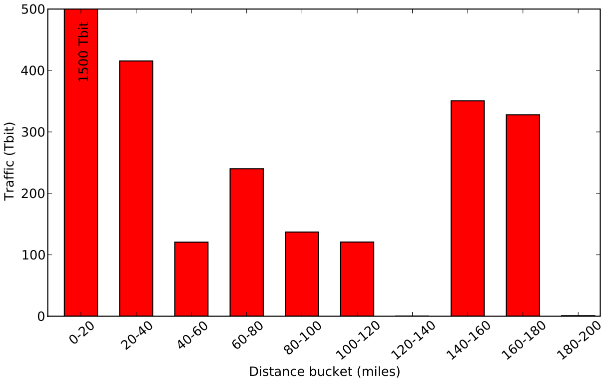

Figure 4: Traffic volume breakdown by distance it traversed.

We explore the potential benefits to an ISP and customers for selling destination-based microcontracts. We first observe that individual traffic flows travel different distances on the ISP backbone, implying that these flows cause the ISP to incur different costs for each flow. We also observe that there are different "types" of customers—some who primarily send traffic regionally, and others who are primarily long-haul.

3.2.1 *Observations*

We first aim to quantify the distance that different traffic flows travel. Figure 4 shows the traffic volume breakdown by distance for traffic carried by a transit ISP in the UK. The distance shown reflects the distance between the PoPs where traffic enters and leaves the network. Most of the traffic stays at the same router or traverses less than 20 miles, but a significant part of the traffic travels more than 100 miles and thus utilizes long distance infrastructure. In fact we observe that 67 clients (13%) send more than half of their traffic over 100 miles and longer distances. This observation implies that an ISP has a potential to improve both its profits and consumer surplus by offering contracts that are more specifically tailored to customers.

3.2.2 *When do destination microcontracts help?*

We aim to quantify the benefits that distance-based microcontracts can provide to ISPs and customers. As before, we assume that customers have a constant number of tasks to accomplish; thus, we keep the aggregate traffic level across all distance intervals constant ($C = \sum_i q_i$). We set cost c_i to be proportional to the distance in the interval i and maximize for profit over all intervals:

$$\begin{aligned} \text{maximize} &: \sum_{i=1}^{n} p_i Q_i(p_i) - \sum_{i=1}^{n} c_i Q_i(p_i) \\ \text{s.t.} &: \sum_{i=1}^{n} Q_i(p_i) = C \\ \text{w.r.t.} &: p_i, \forall i \in \{1 \cdots n\} \end{aligned} \tag{4}$$

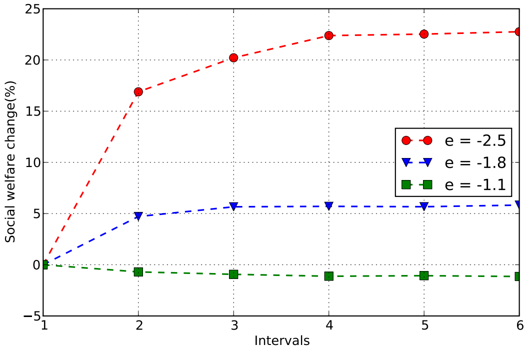

Figure 5: Social welfare change as ISP is increasing the number of pricing tiers. Customer surplus is reduced with more granularities, but the growth in ISP profit outweighs it for higher elasticities.

We transform the objective function and use convex optimization to find a price setting that maximizes ISP profit.

Figure 5 shows how social welfare changes as the ISP offers microcontracts on an increasing number of distance intervals. As expected, as the number of intervals increases, the effect of pricing granularity diminishes. Introducing just three pricing intervals yields about 85% of the social welfare increase achieved with six pricing intervals. In practice, a transit ISP might set one price for traffic staying in the PoP, another price for traffic within a region, and a third price for traffic leaving the continent.

Although not shown in the plot, most of the social welfare gain is due to increasing ISP profits; consumer surplus decreases as the number of intervals increases. When elasticity is low, ISP profit increase is minimal, as it has difficulty encouraging users to switch to shorter distances (since traffic level must remain constant). This trend provide two insights: First, a social planner should encourage adoption of distance microcontracts only if customers in the market are elastic enough. Second, because microcontracts decrease consumer surplus, the consumer will not participate in the market unless ISPs offer somehow share the surplus with customers (*e.g.*, through discounts).

4. Summary and Research Agenda

Although the Internet's applications and users have evolved and diversified over the last decade, the Internet's contract structure has remained stagnant and rigid. This rigid structure has resulted in contracts for Internet connectivity that do not efficiently match user demands for connectivity to the connectivity that ISPs are actually selling. This paper suggests that ISPs might instead sell *microcontracts*, which allow ISPs to "slice" Internet connectivity along various dimensions such as time, destination, volume, and application type. Microcontracts essentially package Internet connectivity according to finer granularities, recognizing that buyers of Internet connectivity may have unique and specific traffic demands (*e.g.*, biased towards some application type, set of destinations, or times of day). Using real traffic demands from a large ISP, we have shown how offering microcontracts whereby customers can purchase connectivity that more closely matches their traffic demands improves both profit for ISPs and utility for buyers.

Judicious application of microcontracts could help solve other real-world problems that operators face today. Destination-based microcontracts between ISPs could help prevent ISPs from having to perform hot-potato routing: the ISP that receives traffic from its neighbor could charge a higher rate for the traffic that will cross its backbone and a lower rate for the traffic destined to its local clients. Application or destination-based microcontracts could allow ISPs to exchange payments for some types of traffic while exchanging other traffic for free.

Although the model we have presented suggests that microcontracts can often benefit both buyers and sellers of Internet connectivity, this paper has merely presented the *case* for microcontracts but has not demonstrated how an ISP can structure their microcontracts to solve problems such as the ones mentioned above. In addition to exploring the feasibility of microcontracts for other dimensions (*i.e.*, volume, application type), our future work will explore how an ISP can realize microcontracts in practice based on their traffic demands. We aim to develop an optimization framework whereby an ISP can determine the price granularities and prices for the contracts that it should sell to each customer, given current traffic demands. We will also explore other practical questions, such as how such microcontracts can be metered or enforced in practice.

REFERENCES

[1] Lat-am operators pushing on-net traffic. `http://www.cellular-news.com/story/32859.php`, 2008.

[2] Interroute internet data transit. `http://www.interoute.com/product_groups/internet_data/internet_services/ip_transit`, 2010.

[3] Arbor Network. Atlas internet observatory. `http://www.arbornetworks.com/`, 2009.

[4] P. Hande, M. Chiang, R. Calderbank, and J. Zhang. Pricing under constraints in access networks: Revenue maximization and congestion management. In *Proc. IEEE INFOCOM*, San Diego, CA, Mar. 2010.

[5] L. Jiang, S. Parekh, and J. Walrand. Time-dependent network pricing and bandwidth trading. In *Proc. IEEE BoD*, 2008.

[6] G. Kesidis, A. Das, and G. D. Veciana. On flat-rate and usage-based pricing for tiered commodity internet services. In *Proc. CISS*, 2008.

[7] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Trans. Netw.*, 8(5):556–567, 2000.

[8] S. Shakkottai, R. Srikant, A. E. Ozdaglar, and D. Acemoglu. The price of simplicity. *IEEE Journal on Selected Areas in Communications*, 26(7):1269–1276, 2008.

[9] A. Technica. AT&T wants to ëducate,čharge heavy data users more. `http://arstechnica.com/gadgets/news/2009/12/att-wants-to-educate-charge-heavy-data-users-more.ars`, 2009.