How Many Tiers? Pricing in the Internet Transit Market

Vytautas Valancius*, Cristian Lumezanu*, Nick Feamster*,
Ramesh Johari†, and Vijay V. Vazirani*
* Georgia Tech † Stanford University

ABSTRACT

ISPs are increasingly selling "tiered" contracts, which offer Internet connectivity to wholesale customers in bundles, at rates based on the cost of the links that the traffic in the bundle is traversing. Although providers have already begun to implement and deploy tiered pricing contracts, little is known about how to structure them. Although contracts that sell connectivity on finer granularities improve market efficiency, they are also more costly for ISPs to implement and more difficult for customers to understand. Our goal is to analyze whether current tiered pricing practices in the wholesale transit market yield optimal profits for ISPs and whether better bundling strategies might exist. In the process, we offer two contributions: (1) we develop a novel way of mapping traffic and topology data to a demand and cost model; and (2) we fit this model on three large real-world networks: an European transit ISP, a content distribution network, and an academic research network, and run counterfactuals to evaluate the effects of different bundling strategies. Our results show that the common ISP practice of structuring tiered contracts according to the cost of carrying the traffic flows (*e.g.*, offering a discount for traffic that is local) can be suboptimal and that dividing contracts based on *both traffic demand and the cost of carrying it* into *only three or four tiers* yields near-optimal profit for the ISP.

Categories and Subject Descriptors

C.2.3 [**Network Operations**]: Network Management

General Terms

Algorithms, Design, Economics

1. INTRODUCTION

The increasing commoditization of Internet transit is changing the landscape of the Internet bandwidth market. Although residential Internet Service Providers (ISPs) and content providers are connecting directly to one another more often, they must still use major Internet transit providers to reach most destinations. These Internet transit customers can often select from among dozens of possible providers [28]. As major ISPs compete with one another, the price of Internet transit continues to plummet: on average, transit prices are falling by about 30% per year [24].

As a result of such competition, ISPs are evolving their business models and selling transit to their customers in many ways to try to retain profits. In particular, many transit ISPs implement pricing strategies where traffic is priced by volume or destination [9]. For example, most transit ISPs offer volume discounts for higher *commit levels* (*e.g.*, customer networks committing to a lower minimum bandwidth receive a higher per-bit price quote than customers committing to a higher minimum bandwidth [24]). Such a market is said to implement *tiered pricing* [13]. Through private communication with network operators, we identified many other instances of tiered pricing already being implemented by ISPs. These pricing instruments involve charging prices on traffic bundles based on various factors, such as how far the traffic is traveling, and whether the traffic is "on net" (*i.e.*, to that ISP's customers) or "off net". Still, we understand very little about the extent to which tiered pricing benefits both ISPs and their customers, or if there might be better ways to structure the tiers. In this paper, we study *destination-based tiered pricing*, with the goal of understanding how ISPs should bundle and price connectivity to maximize their profit.

In this study, we grapple with the balance between the prescriptions of economic theory and a variety of practical constraints and realities. On one hand, economic theory says that higher market granularity leads to increased efficiency [17]. Intuitively, an Internet transit market that prices individual flows is more efficient than one that sells transit in bulk, since customers pay only for the traffic they send, and since ISPs can price those flows according to their cost. On the other hand, various practical constraints prevent Internet transit from being sold in arbitrarily fine granularities. Technical hurdles and additional overhead can make it difficult to implement tiered pricing in current routing protocols and equipment. Additionally, tiered pricing can be more difficult for wholesale customers to understand if there are too many tiers. Transit ISPs would ideally like to come close to maximizing their profit with only a few pricing tiers, since implementing more pricing tiers introduces additional overhead and complexity. Our analysis shows that, indeed, in many cases, an ISP reaps most of the profit possible with infinitesimally fine-grained tiers using only two or three tiers, assuming that those two or three tiers are structured properly.

Although understanding the benefits of different pricing structures is important, modeling them is quite difficult. The model must take as an input existing customer demand and predict how traffic (and, hence, ISP profit) would change in response to pricing strategies. Such a model must capture how customers would respond to any pricing change—for any particular traffic flow—as well as the change in cost of forwarding traffic on various paths in an ISP's network. Of course, many of these input values are difficult to come by even for network operators, but they are especially elusive for researchers; additionally, even if certain values such as costs are known, they change quickly and differ widely across ISPs.

The model we develop allows us to estimate the relative effects of pricing and bundling scenarios, despite the lack of availability of

precise values for many of these parameters. The general approach, which we describe in Section 3, is to start with a demand and cost model and assume both that ISPs are already profit-maximizing and that the current prices reflects both customer demand and the underlying network costs. These assumptions allow us to either fix or solve for many of the unknown parameters and run counterfactuals to evaluate the relative effects of dividing the customer demand into pricing tiers. To drive this model, we use traffic data from three real-world networks: a major international content distribution network with its own network infrastructure; an European transit ISP; and an academic research network. We map the demand and topology data from these networks to a model that reflects the service offerings that real-world ISPs use.

Using our model, we evaluate three scenarios:

- *What happens if an ISP increases the number of tiers for which it sells transit?* We find that profit increases, but the returns diminish as the number of tiers increases: with 3–4 tiers, it is possible to capture 90–95% of the profit that could be captured with an infinite number of granularities, assuming that these tiers are divided in the right way.
- *How do strategies for dividing capacity into distinct bundles and pricing those bundles affect an ISP's profit?* Our analysis shows that ISPs must judiciously choose how they divide traffic into pricing tiers. A naïve approach (*e.g.*, based only on traffic cost or on demand) might require dozens of pricing tiers to capture most of the possible profit. We find that dividing traffic into tiers in a way that accounts for *both traffic demand and the cost of carrying traffic* yields more profit than the current practice that is based only on cost, and is nearly as effective as an optimal division.
- *How do the benefits of various pricing strategies depend on the network topology and traffic demands?* We find that networks with high variability in *cost* of delivering traffic obtain greater benefit from bundling. We also observe that networks with high variability in *demand* require more bundles to capture maximum profit.

We evaluate these and other questions across two customer demand models, four network cost models, and a range of input parameters, such as price sensitivity. Although each of the models might not be perfectly accurate, they yield results that are consistent both across models and with our intuition about markets.

We make three contributions. First, we taxonomize the state of the art and trends in pricing instruments for Internet transit (Section 2). Second, to analyze the effects of tiered pricing, we develop a model that captures demands and costs in the transit market. One of the challenges in developing such a model is applying it to real traffic data, given many unknown parameters (*e.g.*, the cost of various resources, or how users respond to price). Hence, we develop methods for fitting empirical traffic demands to theoretical cost and demand models. We use this approach to evaluate ISP profit for pricing strategies under a range of possible cost models and network topologies (Section 3). Third, we apply our model to real-world traffic matrices and network topologies to characterize a range of simple bundling strategies that are close to optimal (Section 4); we also suggest how these strategies could be implemented in practice (Section 5).

2. BACKGROUND

In this section, we describe the current state of affairs in the Internet transit market. We first taxonomize *what* services (bundles) ISPs are selling. We then provide intuition on *why* ISPs are moving towards tiered wholesale Internet transit service.

2.1 Current Transit Market Offerings

Unfortunately, there is not much public information about the wholesale Internet transit market. ISPs are reluctant to reveal specifics about their business models and pricing strategies to their competitors. Therefore, to obtain most of the information in this section, we engaged in many discussions and email exchanges with network operators. Below, we classify the types of Internet transit service we identified during these conversations. Although much of the information in this section is widely known in the network operations community, it is difficult to find a concise taxonomy of product offerings in the wholesale transit market. The taxonomy below serves as a point of reference for our discussions of tiered pricing in this paper, but it may also be useful for anyone who wishes to better understand the state of the art in pricing strategies in the wholesale transit market.

Transit. Most ISPs offer conventional Internet transit service. Internet transit is sold at a *blended rate*—a single price (usually expressed in $/Mbps/month)—charged for traffic to all destinations. Historically, blended rates have been decreasing by 30% each year [24]. Blended rate is the simplest and yet the most crude way to charge for traffic. If network costs are highly variable, less costly flows in the blended-rate bundle subsidize other, more expensive flows. ISPs often innovate by offering more than one rate: We summarize three pricing models that require two or more rates: (1) *paid peering*, (2) *backplane peering*, and (3) *regional pricing*.

Paid peering is similar to settlement-free peering, except that one network pays to reach the other. A major ISP might separately sell *off-net* routes (wholesale transit) at one rate and *on-net* routes (to reach destinations inside its own network) at another (usually lower) rate. For example, national ISPs in Eastern Europe, Australia, and in other regions may sell local connectivity at a discount to increase demand for local traffic, which is is significantly cheaper than transit to outside global destinations [1]. The on-net routes are also offered at a discount by some major transit ISPs to large content providers, because such transit ISPs can recoup part of the costs from their customers, who congest paid upstream links to transit ISPs by downloading the content. Some instances of paid peering have spawned significant controversy: most recently, Comcast—primarily a network serving end-users—was accused of a network neutrality violation when it forced one tier-1 provider to pay to reach Comcast's customers [16].

Backplane peering occurs when an ISP, in addition to selling global transit through its own backbone, charges a discount rate for the traffic it can offload to its peers at the same Internet exchange. Smaller ISPs buy such a service because they might not meet all the settlement-free peering requirements to peer directly with the ISPs in the exchange. Although many large ISPs discourage this practice, some ISPs deviate by offering backplane peering to retain customers or to maintain traffic ratios with their peers. As with paid peering, the ISP selling backplane peering has to account and charge for at least two traffic flows: one to peers and another to its backbone.

Regional pricing occurs when transit service providers offer different rates to reach different geographic regions. The regions can be defined at different levels of granularity, such as PoP, metro area, regional area, nation, or continent. In some instances, the transit ISP offers access to all regions with different prices; in other instances, the downstream network purchases access only to a specific geographic region (*e.g.*, access only to South America or Australia). In practice, due to the overhead of provisioning and maintaining many sessions to the same customer, ISPs rarely use more than one or two extra price levels for different regions.
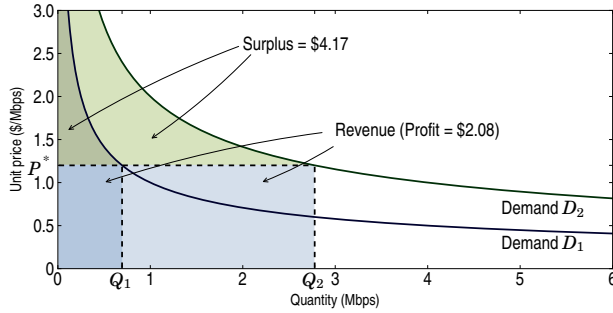
(a) **Blended-rate pricing.** ISP charges a single blended rate P_0.
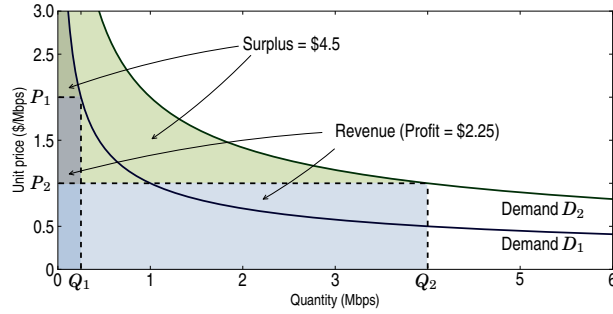
(b) **Tiered pricing.** ISP charges rates P_1 and P_2 for flows.

Figure 1: Market efficiency loss due to coarse bundling.

We speculate that the bundling strategies described above arose primarily from operational and cost considerations. For example, it is relatively easy for a transit ISP to tag which routes are coming from customers and which routes are coming from peers and then in turn sell them separately to its customers. Similarly, it is relatively easy to sell local (*i.e.*, less costly) routes separately. We show that these naïve bundling strategies might not be as effective as bundling strategies that account for both cost and demand.

2.2 The Trend Towards Tiered Pricing

Conventional blended rate pricing is simple to implement, but it may be inefficient. ISPs can lose profit as a result of blended-rate pricing, and customers can lose surplus. This is an example of *market failure*, where goods are not being efficiently allocated between participants of the market. Another outcome of blended-rate pricing is the *increase in direct peering* to circumvent "one-size-fits-all" transit. Both phenomena provide incentives for ISPs to improve their business models to retain revenue. We now explain each of these outcomes.

2.2.1 Profit and surplus loss

Selling transit at a blended rate could reduce profit for transit ISPs and surplus for customers. We define an ISP's *profit* as its revenue minus its costs, and *customer surplus* as customer utility minus the amount it pays to the ISP. Unrealized profit and surplus can occur when ISPs charge a single rate while incurring different costs when delivering traffic to destinations.

Figure 1 illustrates how tiered pricing can increase both the profit for an ISP and the surplus for a customer. The downward-sloping curves represent consumer demand[1] to two destinations. Since

[1]We model consumer demand as *residual demand*. Residual demand accounts for consumption change both due to inherent consumer demand and due to some consumers shifting consumption to substitutes, such as other ISPs (See Section 3.2.1.)
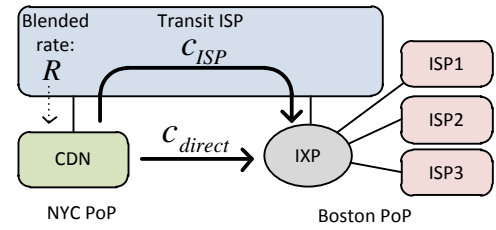

Figure 2: The customer procures a direct link if the cost for such a link is lower than the blended rate $c_{direct} < R$.

the demand slope D_2 is higher than demand slope D_1, the customer has higher demand for the second destination in the ISP's network. Assume that the ISP cost of serving demand D_1 is \$1, while the cost of serving demand D_2 is \$0.5. Modeling demand with constant elasticity (Section 3), the *profit maximizing price* can be shown to be $P_0 = \$1.2/Mbps$. If, however, the ISP is able to offer two bundles, then the profit maximizing prices for such bundles would be $P_1 = \$2.7$ and $P_2 = \$1$. Figure 1(b) shows that this price setup not only increases ISP profit but also increases consumer surplus and thus social welfare.

The market achieves higher efficiency because customers adjust their consumption levels of the ISP network according to their demand and to the prices that the ISP exposes, which directly depend on its costs. Without the ISP's indirect exposure of its costs, the customer consumes less of the cheaper capacity and more of the expensive capacity than it would otherwise. In Section 3, we formalize the market that we have used in this example and propose more complex demand and cost models.

2.2.2 Increase in direct peering

Charging for traffic at a blended rate also provides incentives for client networks to connect directly to geographically close Internet Exchange Points (IXPs). For instance, if a transit ISP charges only blended rate, client network might find geographically close IXPs cheaper to reach by leasing or purchasing private links. While direct peering is generally perceived as a positive phenomenon, for transit ISPs it means less revenue. Direct peering efforts can also diminish economies of scale: instead of using shared ISP infrastructure, customers provision their own connectivity to IXPs.

Figure 2 illustrates an interaction between an upstream ISP and a CDN client (*e.g.*, Google, Microsoft) with its own backbone, which extends to the NYC PoP. The CDN might, or might not, have a content cache at the Boston IXP, but since it does not have its own backbone presence at the IXP, the CDN must pay the upstream ISP to reach it. The ISP offers a blended rate R at the the NYC PoP for all the traffic, including the traffic to the Boston IXP. The blended rate R is set to compensate the upstream provider for the overall traffic mix and, therefore, is *higher* than the amortized cost of most of the cheaper (more localized) flows that ISP is serving (*i.e.*, the flows between the NYC and Boston PoPs). The CDN eventually will procure a direct link to the Boston IXP, if it finds that it can procure such a direct link at an amortized cost $c_{direct} < R$. Assuming the ISP's profit margin is M and flow accounting overhead is A (discussed in Section 5.2), such a direct link presents a *market failure* if $c_{direct} > (M + 1)c_{ISP} + A$, because the customer deploys additional capacity at a higher cost than the ISP could have charged in a tiered market.

Some operators we interviewed confirm that they periodically re-evaluate transit bills and expand their backbone coverage if they find that having own presence in an IXP pays off. In today's transit market, many customers increasingly opt for direct peering [14]; transit service providers are absorbing losses as a result of compet-

itive pressure [24]. Naturally, this pressure increases the incentive for ISPs to adopt a tiered pricing model for local traffic. The central question, then, is how they should go about structuring these tiers. The rest of the paper focuses on this question.

3. MODELING PROFITS, COSTS, AND DEMANDS

We develop demand and cost models that capture ISP profit under various pricing strategies. Although we doubt there is a perfect model for demand in the Internet transit market, we perform our evaluation with two common demand models. Because cost is also difficult to model, we devise four network cost models. We first define ISP profit and then describe demand and cost models.

3.1 ISP Profit

We consider a transit market with multiple ISPs and customers. Each ISP is rational and maximizes its profit, which we express as the difference between its revenue and costs:

$$\Pi(\vec{P}) = \sum_{p_i \in \vec{P}} \left(p_i Q_i(\vec{P}) - c_i Q_i(\vec{P}) \right) \qquad (1)$$

where p_i is the price an ISP sets to deliver flow i, c_i is the unit cost for i, and $Q_i(\vec{P})$ is the demand for i given a vector of prices $\vec{P} = (p_1, p_2, \dots, p_n)$. An ISP chooses the price vector $\vec{P}$ that maximizes its profit. Having a price for each flow allows us to explore different pricing strategies by bundling flows in different ways. For example, blended rate pricing requires p_i to be equal for all i; we can explore different tiered pricing approaches by requiring various subsets of all flows to have the same price.

Given knowledge of both the traffic demand of customers and the costs associated with delivering each flow, we can compute ISP profit. Unfortunately, it is difficult to validate any particular demand function or cost model; even if validation were possible, it is likely that cost structures and customer demand could change or evolve over time. Accordingly, we evaluate ISP profit for various tiered pricing approaches under a variety of demand functions and cost models. Section 3.2 describes the demand functions that we explore, and Section 3.3 describes the cost models that we consider.

3.2 Customer Demand

To compute ISP profit for each pricing scenario, we must understand how customers adjust their traffic demand in response to price changes. We consider two families of demand functions: *constant elasticity* and *logit*.

Constant elasticity demand. The constant elasticity demand (CED) is derived from the well-known *alpha-fair* utility model [21], which is often used to model user utility on the Internet. The alpha-fair utility takes the form of a concave increasing utility function, which emulates a decreasing marginal benefit to additional bandwidth for a user. In this model flow demands are *separable* (*i.e.*, changes in demand or prices for one flow have no effect on demand and prices of other flows). The CED model is most appropriate for scenarios when consumers have no alternatives (*e.g.*, when the content that a customer is trying to reach is not replicated, or the customer needs to communicate with a specific endpoint on the network).

Logit demand. To capture the fact that customers might sometimes have a choice between flows (*e.g.*, sending traffic to alternative destination if the current one becomes too expensive), we also perform our analysis using the *logit model*, where demands are not separable: the price and demand for any flow depend on prices and
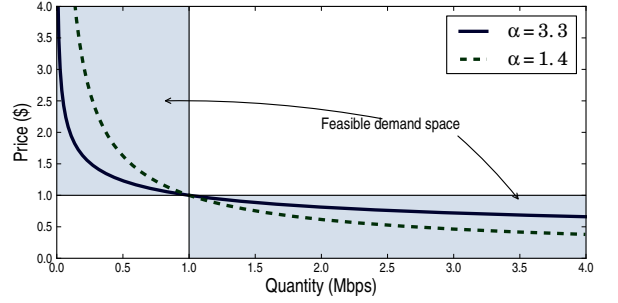

Figure 3: Feasible CED demand functions for $v = 1$.

demands for the other flows. The logit model is frequently used for this purpose in econometric demand estimation [19]. In the logit model, each consumer nominally prefers the flows that offers the highest utility. This matches well with scenarios when consumers have several alternatives (*e.g.*, when requested content is replicated in multiple places).

3.2.1 Constant elasticity demand

The CED demand function is as follows:

$$Q_i(p_i) = \left(\frac{v_i}{p_i} \right)^\alpha \qquad (2)$$

where p_i is the unit price (*e.g.*, \$/Mbit/s), $\alpha \in (1, \infty)$ is the price sensitivity, and $v_i > 0$ is the valuation coefficient of flow i. The demand function can be interpreted to represent either *inherent* consumer demand or *residual* consumer demand, which reflects not only the inherent demand but also the availability of substitutes.

Figure 3 presents example CED demand functions for $v = 1$ and two values of α, 3.3 and 1.4. Higher values of α indicate high elasticity (users reduce use even due to small changes in price). For example, the demand with elasticity $\alpha = 3.3$ might represent the traffic from residential ISPs, who are more sensitive to wholesale Internet prices and who respond to price changes in a more dramatic way. Similarly, the demand with elasticity $\alpha = 1.4$ might represent the traffic from enterprise customers, who are less sensitive to the Internet transit price changes. Although our model does not capture full dynamic interaction between competing ISPs (*e.g.*, price wars), modeling demand as residual allows us to account for the existing competitive environment and switching costs. As discussed above, high elasticity can also indicate that competitors are offering more affordable substitutes, and that switching costs for customers are low. In our evaluation, we use a range of price sensitivity values to measure how ISP profit changes for different values of the elasticity of user demand. The gray area in Figure 3 shows that we can cover all feasible demand functions simply by varying the sensitivity parameter.

CED profit. Using the expressions for ISP profit (Equation 1) and demand (Equation 2), and assuming separability of demand of different flows, the ISP profit is:

$$\Pi(\vec{P}) = \sum_{p_i \in \vec{P}} \left(\frac{v_i}{p_i} \right)^\alpha (p_i - c_i). \qquad (3)$$

CED profit-maximizing price. By differentiating the profit, we find the profit-maximizing price for each flow i:

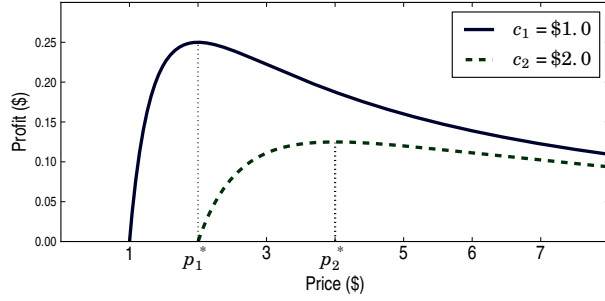$$p_i^* = \frac{\alpha c_i}{\alpha - 1}. \qquad (4)$$

Figure 4: Profit for two flows with identical demand ($v_1 = v_2 = 1.0$, $\alpha = 2$) but different cost.

Figure 4 illustrates profit maximization for two flows that have identical demand functions but different costs. For example, the first flow costs $c_1 = \$1.0$ per unit to deliver and mandates optimal price $p^* = \$2.0$ which results in 0.25 profit. The second flow is more costly thus the profit maximizing price is higher. In this case, the first plot might represent profit for local traffic, while the second plot represents national traffic: ISPs must price national traffic higher than local-area traffic to maximize profit.

CED price for bundled flows. In our evaluation, we test various pricing strategies that bundle multiple flows under the same profit-maximizing price. To find the price for each bundle, we first map real world demands to our model to obtain the valuation v_i and cost c_i for each flow. Then, we differentiate the profit (Equation 3) with respect to the price of each bundle. For example, when we have a single bundle for all flows, we obtain the following profit-maximizing price:

$$P^* = \frac{\alpha \sum_{i=1}^{n} c_i v_i^{\alpha}}{(\alpha - 1) \sum_{i=1}^{n} v_i^{\alpha}} \quad (5)$$

where n is the number of flows. Section 4 details this approach.

3.2.2 Logit demand

The logit demand model assumes that each consumer faces a discrete choice among a set of available goods or services. In the context of data transit, the choice is between different destinations or flows. Following Besanko *et al.* [4], a consumer j using flow i will obtain the utility:

$$u_{ij} = \alpha(v_i - p_i) + \epsilon_{ij}$$

where $\alpha \in (0, \infty)$ is the elasticity parameter, v_i is the "average" consumer's *maximum willingness to pay* for flow i, p_i is a price of using i, and ϵ_{ij} represents consumer j's idiosyncratic preference for i (where ϵ_{ij} has a Gumbel distribution.) The logit model defines the probability that any given consumer will use flow i as a function of the price vector of all flows:

$$s_i(\vec{P}) = \frac{e^{\alpha(v_i - p_i)}}{\sum_j e^{\alpha(v_j - p_j)} + 1} \quad (6)$$

where $\sum_i s_i(\vec{P}) = 1$. The demand for flow i equals the product of $s_i(\vec{P})$ and the total number of consumers (K):

$$Q_i(\vec{P}) = K s_i(\vec{P}). \quad (7)$$

Here, s_i is also called the *market share* of flow i. The model also accounts for the possibility that some customers elect not to send traffic to any destination. The market share for traffic not sent is:
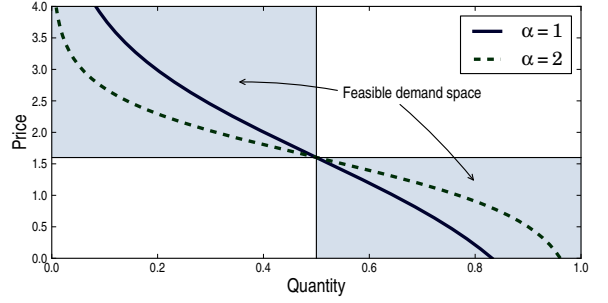

Figure 5: Logit demand function.

$$s_0(\vec{P}) = \frac{1}{\sum_j e^{\alpha(v_j - p_j)} + 1}.$$

Figure 5 shows examples of logit demand functions. We assume a setting with two flows, with two values for the valuation v_i, 1.6 and 1. We fix the price for the first flow to 1, and we vary the price for the second flow between 0 and 4. The figure shows demand curves for the second flow, for two values of α. Similar to the constant elasticity demand model, lower values of α indicate low elasticity of demand, where users need bigger price variations to modify their usage.

Logit profit. Using the expressions for ISP profit (Equation 1) and logit demand (Equation 7), the ISP profit is:

$$\Pi(\vec{P}) = K \sum_{p_i \in \vec{P}} s_i(\vec{P})(p_i - c_i). \quad (8)$$

Logit profit-maximizing prices. To find the profit maximizing price for flow i, we find the first-order conditions for Equation 8:

$$p_i^* = c_i + \frac{1}{\alpha s_0}. \quad (9)$$

Due to the presence of s_0, p_i^* recursively depends on itself and on profit-maximizing prices of other flows. To obtain maximum profit, we develop an iterative heuristic based on gradient descent that starts from a fixed set of prices ($p_i = P_0, \forall i$) and greedily updates them towards the optimum.

Valuation and cost of bundled flows. To test pricing strategies, we first map real traffic demands to the model to find the valuation v_i and cost c_i for each flow i. We then bundle the flows as described in Section 4.2.1. Knowing that $\sum_i s_i = 1$ and applying Equation 6 allows us to compute valuations for any bundle of flows as:

$$v_{bundle} = \frac{\ln\left(\sum_{i=1}^{n} e^{\alpha v_i}\right)}{\alpha} \quad (10)$$

where v_i are valuations of the flows in the bundle. Similarly we can find the average unit cost of combined flows in each bundle:

$$c_{bundle} = \frac{\sum_{i=1}^{n} c_i e^{\alpha v_i}}{\sum_{i=1}^{n} e^{\alpha v_i}}. \quad (11)$$

3.3 ISP Cost

Modeling cost is difficult: ISPs typically do not publish the details of operational costs; even if they did, many of these figures change rapidly and are specific to the ISP, the region, and other
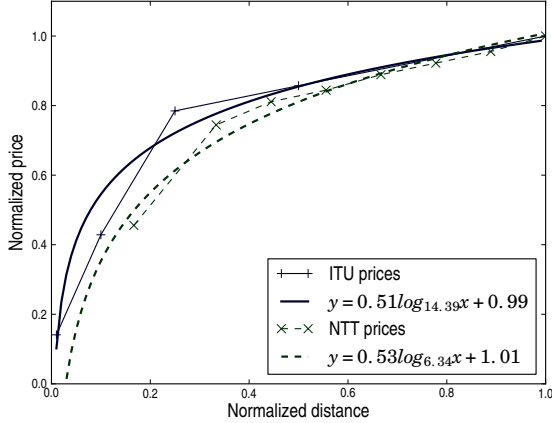
Figure 6: Using price data from ITU [11] and NTT [25] to fit concave distance to cost mapping curve.

factors. To account for these uncertainties, we evaluate our results in the context of several cost models. We also make the following assumptions. First, we assume the more traffic the ISP carries, the higher cost it incurs. Although on a small scale the bandwidth cost is a step function (the capacity is added at discrete increments), on a larger scale we model cost as a linear function of bandwidth. Second, we assume that ISP transit cost changes with distance. Both assumptions are motivated by practice: looking only at specific instances of connectivity, the cost is a step function of distance (*e.g.*, equipment manufacturers sell several classes of optical transceivers, where each more powerful transceiver able to reach longer distances costs progressively more than less powerful transceivers [7]). Over a large set of links, we can model cost as a smooth function of distance.

The cost models below offer only *relative* flow-cost valuations (*e.g.*, flow A is twice as costly as flow B); they do not operate on absolute costs. These relative costs must be reconciled with the blended prices used to derive customer valuations. We describe methods for reconciling these values in Section 4.1. Each cost model has a generic tuning parameter, denoted as θ, which we use in the evaluation.

Linear function of distance. The most straightforward way to model ISP's costs as a function of distance is to assume cost increases linearly with distance. Although, in some cases, this model does not hold (*e.g.*, crossing a mountain range is more expensive than crossing a region with flat terrain), we often observe that ISPs charge linearly in the distance of communication [5, 6]. As we model cost as linear function of distance, we set cost $c_i = \gamma d_i + \beta$, where γ is a scaling coefficient that translates from relative to real costs, β is a base cost (*i.e.*, the fixed cost that the ISP incurs for communicating over any distance), and d_i is the geographical distance between the source and destination served by an ISP. We describe how we determine γ in Section 4.1. We model the base cost β as a fraction of the maximum cost without the base component. More formally: $\beta = \theta \max_{j \in 1 \dots n} \gamma d_j$, where θ in this cost model is a relative base cost fraction, and n is number of flows with different cost. For example, given distances 1, 10, and 100 miles, $\gamma = \$1/mile$, and $\theta = 0.1$, the resulting base cost β is $10, and thus flow costs are $11, $20, and $110. In the evaluation, we vary θ to observe the effects of different base costs. For example, low θ values (low base cost) here represent a case where link distance is the largest contributor to the total cost.

Concave function of distance. We are also aware of ISPs that price transit as a concave function of distance [11, 25]. For this scenario we model the ISP's cost as $c_i = \gamma(a \log_b d_i + c) + \beta$. Figure 6 shows a concave curve fitting to two price data sets, resulting in $a \approx 0.5$, $b \approx 6$, and $c \approx 1$ for normalized prices and distance. As in the case of linear cost, we set the base offset cost $\beta = \theta \max_{j \in 1 \dots n} \gamma(a \log_b d_i + c)$. We use θ in the evaluation to change link distance contribution to the total cost.

Function of destination region. As described in Section 2.2, both private communication with network operators and publicly available data suggest that ISPs can also charge for traffic at rates depending on the region where traffic is destined [8,27,31]. For example, an ISP might have less expensive capacity in the metropolitan area than in the region, less expensive capacity in the region than in the nation, and less expensive capacity in the nation than across continental boundaries. We divide flows into three categories: *metropolitan*, *national*, and *international*. We map the flows into these categories by using data from the GeoIP [18] database: flows that originate and terminate in the same city are classified as metro, and flows that start and end in the same country are classified as national; all other flows are classified as international. For EU ISP we only have distances between traffic entry and exit points, thus we classify flows traveling less than 10 miles as metro, flows that travel less than 100 miles as national, and longer flows as international. We set the costs as follows: $c_{metro} = \gamma$, $c_{nation} = \gamma 2^\theta$, and $c_{int} = \gamma 3^\theta$. This form allows us to test scenarios when there is no cost difference between regions ($\theta = 0$), the cost differences are linear ($\theta = 1$), and costs are different by magnitudes ($\theta > 1$).

Function of destination type. As we described in Section 2.2, ISPs offer discounts for the traffic destined to their customers ("on net" traffic), while charging higher rates for traffic destined for their peers ("off net" traffic). These offerings are motivated by the fact that ISPs do recover some of their transport cost for the traffic sent to other customers. In our evaluation, we model this cost difference by setting the cost of the traffic to peers to be twice as costly than traffic to other customers. The logic behind such a model is that when an ISP sends the traffic between two customers, it gets paid twice by both customers, but when an ISP sends traffic between a customer and a peer it is only paid by the customer. The parameter θ indicates a fraction of traffic at each distance that is destined to clients, as opposed to traffic that is destined to peers and providers.

4. EVALUATING TIERED PRICING STRATEGIES

In this section, we evaluate the efficiency of destination-based tiered pricing using the model presented in Section 3 and real topology and demand data from large networks. Our goal is to understand how the profit that an ISP extracts from offering tiered-pricing depends on the number of tiers (or bundles), the bundling strategy used, and the network topology and traffic demand.

One of the major challenges we face is that we cannot know some aspects of the cost and demand models, or even which model to use. We use the ISP data to derive model parameters, such as valuation or cost, and evaluate the profit of each strategy across models and input parameters. Figure 7 presents an overview of our approach for computing ISP profits.

Our evaluation yields several important results. First, we show that an ISP needs only 3–4 bundles to capture 90–95% of the profit provided by an infinite number of bundles, if it bundles the traffic appropriately. Second, choosing a bundling strategy that considers both flow demand and cost is almost as effective as an exhaustive search for the best combination of bundles. Finally, we observe

Figure 7: We evaluate the effect of tiered pricing on Internet transit by separately modeling the demand and cost of traffic and the way ISPs bundle flows under the same price. At each step we use real-world data to derive unknown parameters.

Data set	Date	Distance (miles)		Traffic (Gbps)	
		w-avg	CV	Aggregate	CV
EU ISP	11/12/09	54	0.70	37	1.71
CDN	12/02/09	1988	0.59	96	2.28
Internet 2	12/02/09	660	0.54	4	4.53

Table 1: Data sets used in our evaluation. The columns represent: the network, data capture date, demand-weighted average of flow distances, coefficient of variation (CV) of flow distances, aggregate traffic per second, and CV of demand of different flows.

that the topology and traffic of a network influences its bundling strategies: networks with higher coefficient of variation of demand need more bundles to extract maximum profit.

4.1 Mapping Data to Models

Because we do not know the parameters that we need to compute the profit-maximizing prices and the maximum profit for each bundling strategy, we must derive them. We first describe the data and how we extract the necessary information for computing model parameters. Then, we show how to apply the demand models to the real traffic demands to compute the valuation coefficients v_i for each flow i. Finally, we derive the ISP's cost for servicing the flow by applying flow distance information to each of the cost models.

4.1.1 Data Sources

We use demand and topology data from three networks: a European ISP serving thousands of business customers (EU ISP), one of the largest CDN providers in the world (CDN), and a major research network in United States (Internet 2). The data consists of sampled NetFlow records from core routers in each network for 24 hours. Table 1 presents more details about the data sets.

To drive our model, we must compute the traffic volume (which captures consumer demand) and the distance between the source and destination of each flow (which captures the relative cost of transit). To do so, we extract the source and destination IP and port information, as well as the traffic level, from each NetFlow record. We obtain the demand for each flow by aggregating all records of the flow, while ensuring that we do not double-count records that are duplicated on different routers.

To compute distances that reflect the ISP's cost of sending traffic, we use the following heuristics. For the EU ISP, the distance that each flow travels in the ISP's network is the geographical distance

between the flow's entry and exit points, whose identity and location is known. For the CDN, we use the GeoIP database [18] to estimate the distance to the destination. Although this may not reflect the real distance that a packet travels (because part of the path may be covered by another ISP), we assume that it is still reflective of the cost incurred by the CDN. Finally, for Internet2, because each flow may traverse multiple routers, we use the port information to identify the links the flow has traversed. The distance each flow traverses is the sum of the links in the path, where the link length is the geographical distance between the neighboring routers.

4.1.2 Discovering valuation coefficients

The valuation coefficient v_i indicates the valuation of flow i. We need the valuation coefficeint v_i to capture how the demand for flow i varies with price (Equations 2 and 7) and thus affect the ISP profit. To find v_i for each flow, we assume that ISPs charge the same blended price P_0 for each flow and map the observed traffic demand from the data to each demand model.

CED valuation coefficient. From Equation 2 we obtain:

$$v_i = \frac{q_i^{\left(\frac{1}{\alpha}\right)}}{P_0}$$

where q_i is observed demand on flow i and α is the sensitivity coefficient that we vary in the evaluation.

Logit demand valuation coefficient. Starting with Equation 7 and knowing that $s_0 + \sum s_i = 1$ and $s_0 = \frac{1}{\sum e^{\alpha(v_j - P_0)} + 1}$, we obtain:

$$v_i = \frac{\log s_i - \log s_0}{\alpha} + P_0$$

where s_i is the market share of flow i. We vary s_0 in the evaluation and compute the remaining market shares from observed traffic as $s_i = \frac{q_i(1 - s_0)}{\sum q_i}$.

4.1.3 Discovering costs

To estimate the ISP profit, we must know the cost that an ISP incurs to service each flow. However, the data provides information only about the distance each flow traverses, which reflects only the relative cost (*e.g.*, flow A is twice as costly as flow B), rather than an absolute cost value. To normalize the cost of carrying traffic to the same units as the price for the flow, we introduce a scaling parameter γ, where $c_i = \gamma f(d_i)$ and d_i is the distance covered by i (see Section 3.3). For each demand model, we compute the scaling parameter by assuming (as in computing valuation coefficients) that ISPs are rational and profit maximizing, and charge the same price P_0 for each flow.

CED. Using Equation 5 and substituting c_i for $\gamma f(d_i)$, we find:

$$\gamma = \frac{P_0(\alpha - 1) \sum_{i=1}^{n} v_i^{\alpha}}{\alpha \sum_{i=1}^{n} f(d_i) v_i^{\alpha}}$$

Logit demand. Differentiating profit Equation 8 and substituting c_i for $\gamma f(d_i)$, we can express γ:

$$\gamma = \frac{\sum \left(e^{\alpha(v_i - P_0)} \left(\alpha P_0 - 1 - \sum_{i=1}^{n} e^{\alpha(v_i - P_0)} \right) \right)}{\alpha \sum_{i=1}^{n} f(d_i) e^{\alpha(v_i - P_0)}}$$

4.2 How Should Tiers Be Structured?

ISPs must judiciously choose how they bundle traffic flows into tiers. As shown in Section 2.1, today's ISPs often offer at most two or three bundles with different prices. We define six bundling strategies that classify and group traffic flows according to their cost, demand, or potential profit to the ISP. We then evaluate them
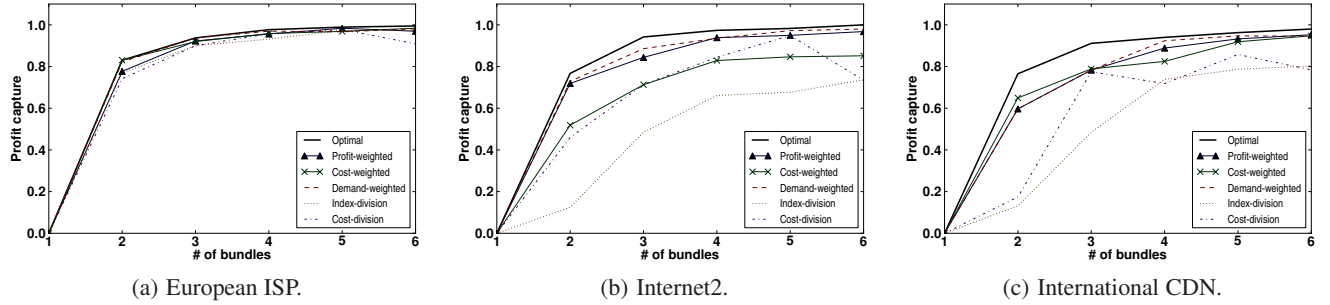
(a) European ISP. (b) Internet2. (c) International CDN.

Figure 8: Profit capture for different bundling strategies in constant elasticity demand.

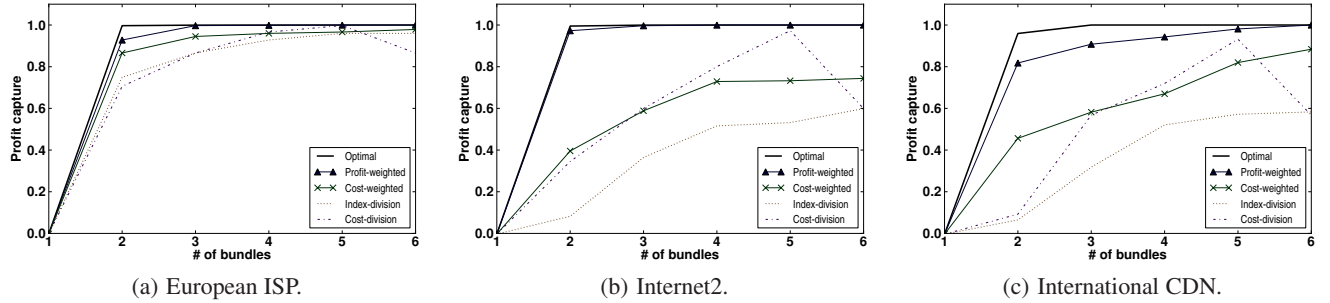(a) European ISP. (b) Internet2. (c) International CDN.

Figure 9: Profit capture for different bundling strategies in logit demand.

and show that, assuming the right bundling strategy is used, ISPs typically need only a few bundles to collect near-optimal profit.

4.2.1 Bundling strategies

Optimal. We exhaustively search all possible combinations of bundles to find the one that yields the most profit. This approach gives optimal results and also serves as our *baseline* against which we compare other strategies. Computing the optimal bundling is computationally expensive: for example, there is more than a billion ways to divide one hundred traffic flows into six pricing bundles. Presented below, all of the other bundling strategies employ heuristics to make bundling computationally tractable.

Demand-weighted. In this strategy, we use an algorithm inspired by token buckets to group traffic flows to bundles. First, we set the overall token budget as the sum of the original demand of all flows: $T = \sum_i q_i$. Then, for each bundle j we assign the same token budget $t_j = T/B$, where B is the number of bundles we want to create. We sort the flows in decreasing order of their demand and traverse them one-by-one. When traversing flow i, we assign it to the first bundle j that either has no flows assigned to it or has a budget $t_j > 0$. We reduce the budget of that bundle by q_i. If the resulting budget $t_j < 0$, we set $t_{j+1} = t_{j+1} + t_j$. After traversing all the flows, the token budget of every bundle will be zero, and each flow will be assigned to a bundle. The algorithm leads to separate bundles for high demand flows and shared bundles for low demand flows. For example, if we need to divide four flows with demands 30, 10, 10, and 10 into two bundles, the algorithm will place the first flow in the first bundle, and the other three flows in the second bundle.

Cost-weighted. We use the same approach as in demand-weighted bundling, but we set the token budget to $T = \sum_i 1/c_i$. When placing a flow in a bundle we remove a number of tokens equal to the inverse of its cost. This approach creates separate bundles for local flows and shared bundles for flows traversing longer distances. The

current ISP practices of offering *regional pricing* and *backplane peering* maps closely to using just two or three bundles arranged using this cost-weighted strategy.

Profit-weighted. The bundling algorithms described above consider cost and demand separately. To account for cost and demand together, we estimate *potential profit* each flow could bring. We use the potential profit metric to apply the same weighting algorithm as in cost and demand-weighted bundling. In case of constant elasticity demand, we derive potential profit of each flow i:

$$\pi_i = \frac{v_i^\alpha}{\alpha} \left(\frac{\alpha c_i}{\alpha - 1} \right)^{1-\alpha} \qquad (12)$$

For the logit demand, substituting p_i in Equation 9 yields:

$$\pi_i = K s_i (p_i - c_i) = \frac{K s_i}{\alpha s_0} \propto q_i \qquad (13)$$

Cost division. We find the most expensive flow and divide the cost into ranges according to that value. For example, if we want to introduce two bundles and the most expensive flow costs $10/Mbps/month to reach, we assign flows that cost $0–$4.99 to the first bundle and flows that cost $5–$10 to the second bundle.

Index division. Index-division bundling is similar to cost division bundling, except that we rank flows according to their cost and use the rank, rather than the cost, to perform the division into bundles.

4.2.2 The effects of different bundling strategies

To evaluate the bundling strategies described above, we compute the profit-maximizing prices and measure the resulting pricing outcome in terms of *profit capture*. Profit capture indicates what fraction of the maximum possible profit—the profit attained using an infinite number of bundles—the strategy captures. For example, if the maximum attainable profit is 30% higher than the original profit, while the profit from using two bundles is 15% higher than the original profit, the profit capture with two
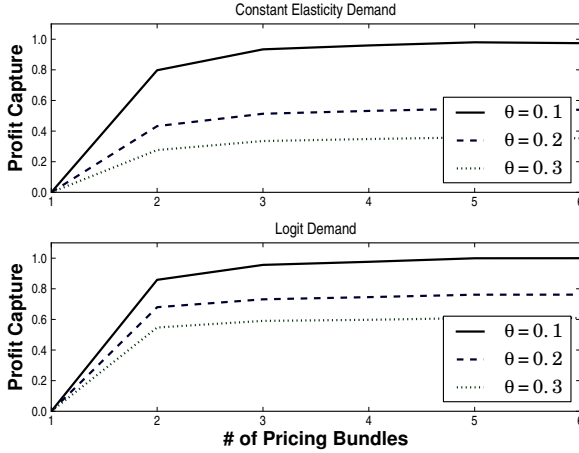
Figure 10: Profit increase in EU ISP network using linear cost model.

bundles attains 0.5 of profit capture. Formally, profit capture is $(\pi_{\text{new}} - \pi_{\text{original}})/(\pi_{\text{max}} - \pi_{\text{original}})$.

Figures 8 and 9 show the profit capture for different bundling strategies, across the three data sets, while varying the number of bundles. For the results shown here, we use both the constant elasticity and the logit demand models and the linear cost model. We set the price sensitivity α to 1.1, the original, blended rate P_0 to $20, the cost tuning parameter θ to 0.2, and the original market fraction that sends no traffic s_0 to 0.2. We explore the effect of varying these parameters in Section 4.3.

Optimal versus heuristics-based bundling. With an appropriate bundling strategy, the ISP attains maximum profit with just 3–4 bundles. As expected, the optimal flow bundling strategy captures the most profit for a given number of bundles. We observe that the EU ISP captures more profit with two bundles than other networks. We attribute this effect to the low coefficient of variation (CV) of demand to different destinations, which limits the benefits of having more pricing bundles. We also discover that, given fixed demand, a high CV of distance (cost) leads to higher absolute profits. With only minor exceptions, the profit-weighted bundling heuristic is almost as good as the the optimal bundling, followed by the cost-weighted bundling heuristic. Deeper analysis, beyond the scope of this work, could show what specific input data conditions cause the profit-weighted flow bundling heuristic to produce bundlings superior to the cost-weighted heuristic.

Logit profit capture. Maximum profit capture occurs more quickly in the logit model because (1) the total demand (including s_0 option) is constant, and (2) the model is sensitive to differences in valuation of different flows. When there is a flow with a significantly higher difference between valuation and cost $(v_i - c_i)$, it absorbs most of the demand. In this model, with just two pricing tiers, local and non-local traffic are separated into distinct bundles that closely represent the *backplane peering* and *regional pricing* for local area service models.

4.3 Sensitivity Analysis

We explore the robustness of our results to cost models and input parameter settings. As we vary an input parameter under test, other parameters remain constant. Unless otherwise noted, we use profit-weighted bundling, the EU ISP dataset, sensitivity $\alpha = 1.1$, the linear cost model with base cost $\theta = 0.2$, blended rate $P_0 = \$20.0$,
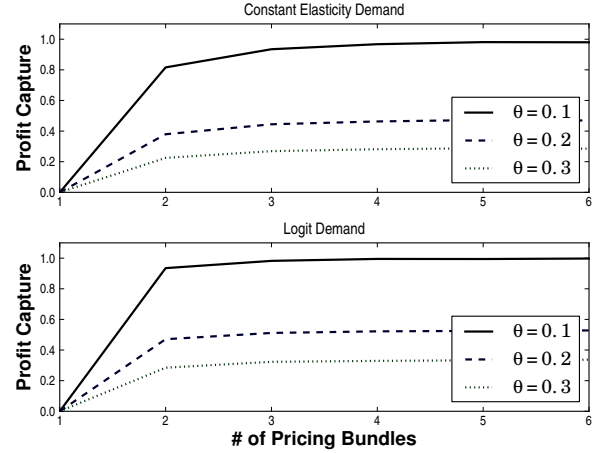

Figure 11: Profit increase in EU ISP network using concave cost model.

and, in the logit model, $s_0 = 0.2$ (the original market fraction that sends no traffic).

4.3.1 Effects of cost models

We aim to see how cost models and settings within these models qualitatively affect our results from the previous section. We show how profit changes as we increase the number of bundles for different settings of the cost model parameters (θ), described in Section 3.3. We find that for different θ settings most of the attainable profit is still captured in 2-3 bundles. Unlike in other sections, in Figures 10–13, we normalize the profit of all the plots in the graphs to the highest observed profit. In other words, π_{max} in these figures is not the maximum profit of each plot, but the maximum profit of the plot with highest profit in the figure. Normalizing by the highest observed profit allows us to show how changing the parameter θ affects the amount of profit that the ISP can capture.

Linear cost. Figure 10 shows profit increase in the EU ISP network as we vary the number of bundles for different settings of θ. As expected, most of the profit is still attained with 2–3 pricing bundles. We also observe that the increase in the base cost (θ) causes a decline in the maximum attainable profit. The reduction in maximum attainable profit is expected, as increasing the base cost reduces the coefficient of variation (CV) of the cost of different flows and thus reduces the opportunities for variable pricing and profit capture. We can also see, as shown in previous section, that the logit demand model attains more profit than the constant elasticity demand model with the same number of pricing bundles.

Concave cost. Figure 11 shows the profit increase as we vary the number of bundles for different settings of θ for the concave cost model. The observations and results are similar to the linear cost model, with one notable exception. The amount of profit the ISP can capture decreases more quickly in the concave cost model than in the linear cost model for the same change in the base-cost parameter θ. This is due to the lower CV of cost in the concave model than in the linear cost model. In other words, applying the log function on distance (as described in Section 3.3) reduces the relative cost difference between flows traveling to local and remote destinations.

Regional cost. In the regional cost model, the parameter θ is an exponent which adjusts the price difference between three different regions: local, national, and international. Figure 12 shows the profit increase in the EU ISP network as we vary number of bundles for different settings of θ. Higher θ values result in a higher CV of
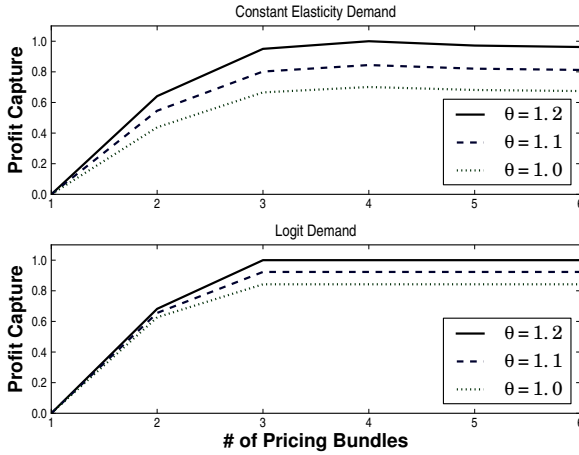
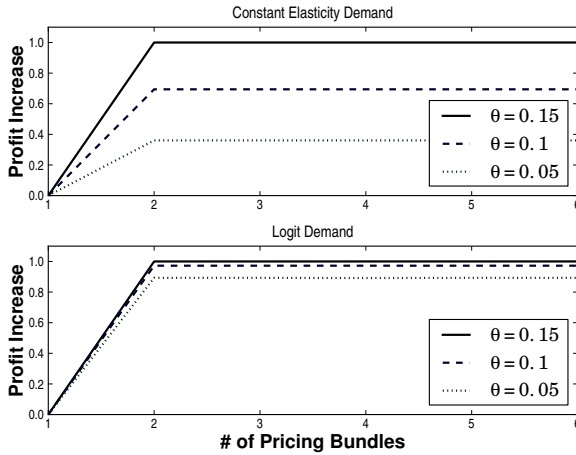Figure 12: Profit increase in the EU ISP network using regional cost model.

Figure 13: Profit increase in the EU ISP network using destination type cost model.

Figure 14: Minimum profit capture for a fixed number of bundles over a range of α between 1 and 10.

Figure 15: Minimum profit capture for a fixed number of bundles over a range of starting prices $P_0 \in [5, 30]$.

cost in different regions which, in turn, in both demand models produces higher profit. Using constant elasticity demand we observe a small dip in profit when using five and six bundles, which recovers later with more bundles. Such dips are expected when there are only a few traffic classes. For example, if traffic had just two distinct cost classes, two judiciously selected bundles could capture most of the profit. Adding a third bundle can reduce the profit if that third bundle contains flows from both of the classes (as may happen in a suboptimal bundling).

Destination type-based cost. Destination type-based cost model emulates "on-net" and "off-net" types of traffic in an ISP network. As described in Section 3.3, we assume that "on-net" traffic costs less than "off-net" traffic. We vary θ, which represents a fraction of "on-net" traffic in each flow. The standard profit-weighting algorithm does not work well with the destination type-based cost model. The effect observed in the regional cost model—where five bundles produce slightly lower profit then four bundles—is more pronounced when we have just two distinct flow classes. One heuristic that works reasonably well is as follows: we update the profit-weighting heuristic to never group traffic from two different classes into the same bundle. Figure 13 shows how profit increases
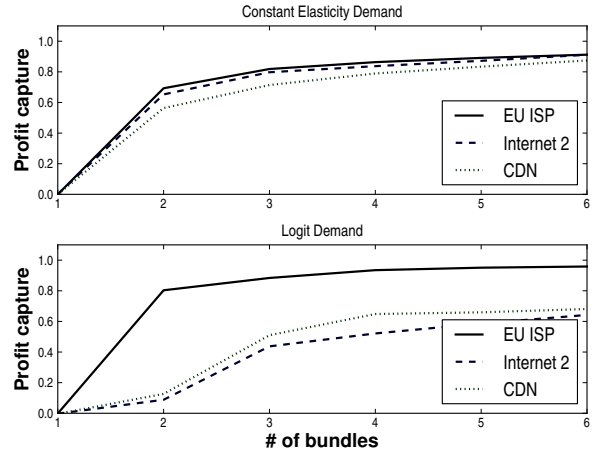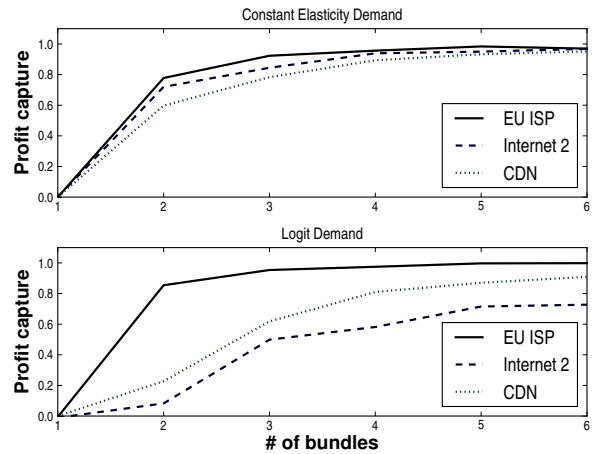
with an increasing number of bundles. Since there are two major classes of traffic ("on-" and "off-net"), most profit is attained with two bundles for both demand models. In this cost model, as in other cost models, the same change in CV of cost (induced by the parameter θ) causes a greater change in profit capture for constant elasticity demand than for logit demand.

4.3.2 Sensitivity to parameter settings

The models we use rely on a set of parameters, such as price sensitivity (α), price of the original bundle (P_0), and, in the logit model, the share of the market that corresponds to deciding not to purchase bandwidth (s_0). In this section, we analyze how sensitive the model is to the choice of these parameters.

Figures 14–16 show how profit capture is affected by varying price sensitivity α, blended rate P_0, and non-buying market share s_0, respectively. Each data point in the figures is obtained by varying each parameter over a range of values. We vary α between 1 and 10, P_0 between 5 and 30, and s_0 between 0 and 0.9. As we vary the parameters, we select and plot the *minimum* observed profit capture over the whole parameter range, for the profit-weighted strategy with different numbers of bundles. In other words, these plots
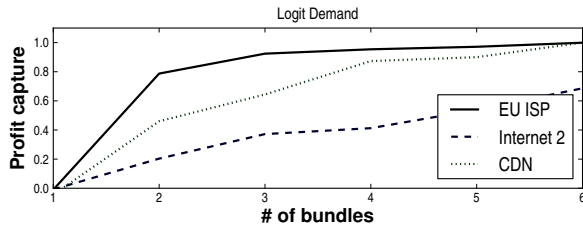
Figure 16: Maximum profit capture for a fixed number of bundles over a range of fractions of users who decide not to participate in the market $s_0 \in (0, 1)$.

show the worst case relative profit capture for the ISP over a range of parameter values. The trend of these minimum profit capture points is qualitatively similar to patterns in Figures 8 and 9. For example, using the CED model and grouping flows in two bundles in the EU ISP yields around 0.8 profit capture, regardless of price sensitivity, blending rate, and market share. These results indicate that our model is robust to a wide range of parameter values.

5. IMPLEMENTING TIERED PRICING

ISPs can implement the type of tiered pricing that we describe in Section 4 without any changes to their existing protocols or infrastructure, and ISPs may already be using the techniques we describe below. If that is the case, they could simply apply a profit-weighted bundling strategy to re-factor their pricing to improve their profit, possibly without even making many changes to the network configuration. We describe two tasks associated with tiered pricing: associating flows with tiers and accounting for the amount of traffic the customer sends in each tier.

5.1 Associating Flows with Tiers

Associating each flow (or destination) with a tier can be done within the context of today's routing protocols. When the upstream ISP sends routes to its customer, it can "tag" routes it announces with a label that indicates which tier the route should be associated with; ISPs can use BGP extended communities to perform this tagging. Because the communities propagate with the route, the customer can establish routing policies on every router within its own network based on these tags.

Suppose that a large transit service provider has routers in different geographic regions. Routers at an exchange point in, say, New York, might advertise routes that it learned in Europe with a special tag indicating that the path the route takes is trans-Atlantic and, hence, bears a higher price than other, regional routes. The customer can then use the tag to make routing decisions. For example, if a route is tagged as an expensive long-distance route, the customer might choose to use its own backbone to get closer to destination instead of performing the default "hot-potato" routing (*i.e.*, offloading the traffic to a transit network as quickly as possible). A large customer might also use this pricing information to better plan its own network growth.
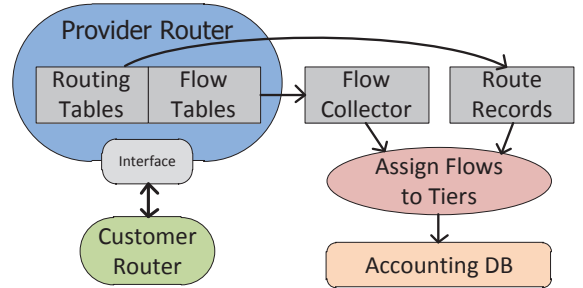
5.2 Accounting

Implementing tiered pricing requires accounting for traffic either on a *per-link* or *per-flow* basis.

Link-Based Accounting. As shown in Figure 17(a), an edge router can establish two or more physical or virtual links to the customer, with a Border Gateway Protocol (BGP) [29] session for each physical or virtual link. In this setup, each pricing tier would have a separate link. Each link carries the traffic only to the set of destinations
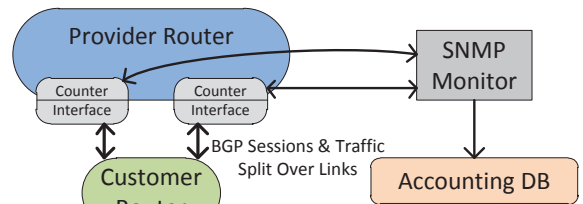
(a) SNMP-Based Accounting.

(b) Flow-Based Accounting.

Figure 17: Implementing accounting for tiered pricing.

advertised over that session (*e.g.*, on-net traffic, backplane peering traffic). Because each link has a separate routing session and only exchanges routes associated with that pricing tier, the customer and provider can ensure that traffic for each tier flows over the appropriate link: The customer knows exactly which traffic falls into which pricing tier based on the session onto which it sends traffic. Billing may also be simpler and easier to understand, since, in this mode, a provider can simply bill each link at a different rate. Unfortunately, the overhead of this accounting method grows significantly with the number of pricing levels ISP intends to support.

Flow-Based Accounting. In flow-based accounting, as in traditional peering and transit, an upstream ISP and a customer establish a link with a single routing session. As shown in Figure 17(b), the accounting system collects both flow statistics (*e.g.*, using Net-Flow [23]) and routing information to determine resource usage. For the purposes of accounting, bundling effectively occurs after the fact: flows can be mapped to distances using the routing table information and priced accordingly, exactly as we did in our evaluation in Section 4. Assuming flow and routing information collection infrastructure in place, flow-based accounting may be easier to manage, and it is easier to bundle flows into different bins according to various bundling strategies (*e.g.*, profit-weighted bundling) *post facto*.

6. RELATED WORK

Developing and analyzing pricing models for the Internet is well-researched in both networking and economics. Two aspects are most relevant for our work: the unbundling of connectivity and the dimensions along which to unbundle it. Although similar studies of pricing exist, none have been evaluated in the context of real network demand and topology data.

The unbundling of connectivity services refers to the setting of different prices for such services along various usage dimensions such as volume, time, destination, or application type. Seminal works by Arrow and Debreu [2] and McKenzie [20] show that markets where commodities are sold at infinitely small granularities are more efficient. More recent studies however, demonstrate that unbundling may be inefficient in certain settings, such as when selling

information goods with zero or very low marginal cost (such as access to online information) [3, 15, 22]. This is not always the case with the connectivity market, where ISPs incur different costs to deliver traffic to different destinations. In addition, many service providers already use price discrimination [26].

Kesidis *et al.* [13] and Shakkottai *et al.* [30] study the benefits of pricing connectivity based on volume usage and argue that, with price differentiation, one can use resources more efficiently. In particular, Kesidis *et al.* show that usage-based unbundling may be even more beneficial to access networks rather than core networks. Time is another dimension along which providers can unbundle connectivity. Jiang *et al.* [12] study the role of time preference in network prices and show analytically that service providers can achieve maximum revenue and social welfare if they differentiate prices across users and time. Hande *et al.* [10] characterize the economic loss due to ISP inability or unwillingness to price broadband access based on time of day.

7. CONCLUSION

As the price of Internet transit drops, transit providers are selling connectivity using "tiered" contracts based on traffic cost, volume, or destination to maintain profits. We have studied two questions: How does tiered pricing benefit both ISPs and their customers? and How should ISPs structure the connectivity tiers they sell to maximize their profits? We developed a model for an Internet transit market that helps ISPs evaluate how they should arrange traffic into different tiers, and how they should set prices for each of those tiers. We have applied our model to traffic demand and topology data from three large ISPs to evaluate various bundling strategies.

We find that the common ISP practice of structuring tiered contracts according to the cost of carrying the traffic flows (*e.g.*, offering a discount for traffic that is local) is suboptimal. Dividing the contract into only three or four tiers based on *both traffic cost and demand* yields near-optimal profit for the ISP; other strategies such as cost division bundling also work well. We also find that networks with primarily lower cost traffic (either local or traveling short distances) require fewer tiers to extract maximum profit than other networks do.

Acknowledgments

REFERENCES

[1] Adam Internet. Unmetered content. `http://www.adam.com. au/unmetered/unmetered.php`. Retrieved: June 2011.

[2] K. J. Arrow and G. Debreu. Existence of equilibrium for competitive economy. *Econometrica*, 22(3):265–290, 1954.

[3] Y. Bakos and E. Bronjolfsson. Bundling and competition on the Internet. *Marketing Science*, 19(1), Jan. 1998.

[4] D. Besanko, S. Gupta, and D. Jain. Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science*, 44:1533–1547, November 1998.

[5] BNSL. Tariff for leased lines. `http://www.bsnl.co.in/ service/2mbps.pdf`. Retrieved: June 2011.

[6] Chunghwa Telecom. Leased line pricelist. `http://www.cht. com.tw/CHTFinalE/Web/Business.php?CatID=476`. Retrieved: June 2011.

[7] Cisco. SFP optics for gigabit ethernet applications. `http://www. cisco.com/en/US/prod/collateral/modules/ ps5455/ps6577/product_data_ sheet0900aecd8033f885.html`. Retrieved: June 2011.

[8] Etisalat ISP. Leased circuit rental charges. `http://tinyurl. com/66tfvj6`. Retrieved: June 2011.

[9] Guavus. Profitable tiered pricing. `http://www.guavus.com/ solutions/tiered-pricing`. Retrieved: June 2011.

[10] 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.

[11] ITU Telecommunication Indicator Handbook. `http://www.itu. int/ITU-D/ict/publications/world/material/ handbook.html`. Retrieved: June 2011.

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

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

[14] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. In *Proc. ACM SIGCOMM*, New Delhi, India, Aug. 2010.

[15] J.-J. Laffont, S. Marcus, P. Rey, and J. Tirole. Internet interconnection and the off-net-cost principle. *The Rand Journal of Economics*, 34(2), 2003.

[16] Level 3 Communications. Level 3 statement concerning Comcast's actions. `http://www.level3.com/index.cfm? pageID=491&PR=962`. Retrieved: June 2011.

[17] A. Mas-Colell, M. Whinston, J. Green, and U. P. F. F. de Ciencies Econnmiques i Empresarials. *Microeconomic theory*, volume 981. Oxford university press New York, 1995.

[18] MaxMind GeoIP Country. `http://www.maxmind.com/app/ geolitecountry`. Retrieved: June 2011.

[19] D. McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers Of Econometrics*, 1974.

[20] L. W. McKenzie. On the existence of general equilibrium for a competitive economy. *Econometrica*, 1959.

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

[22] P. Nabipay, A. M. Odlyzko, and Z.-L. Zhang. Flat versus metered rates, bundling, and "bandwidth hogs". In *6th Workshop on the Economics of Networks, Systems, and Computation*, San Jose, CA, June 2011.

[23] Cisco NetFlow. `http://www.cisco.com/en/US/ products/ps6601/products_ios_protocol_group_ home.html`. Retrieved: June 2011.

[24] W. Norton. DrPeering.net. `http://drpeering.net`.

[25] NTT East. Leased circuit price list. `http://www.ntt-east. co.jp/senyo_e/charge/digital.html`. URL retrieved January 2011.

[26] A. M. Odlyzko. Network neutrality, search neutrality, and the never-ending conflict between efficiency and fairness in markets. *Review of Network Economics*, 8(1):40–60, Mar. 2009.

[27] ORE. Wholesale leased lines price list. `http://www. otewholesale.gr/Portals/0/LEASED%20LINES_ Pricelist_ENG_081110.pdf`. Retrieved: June 2011.

[28] Peering Database. `http://www.peeringdb.com`.

[29] Y. Rekhter, T. Li, and S. Hares. *A Border Gateway Protocol 4 (BGP-4)*. Internet Engineering Task Force, Jan. 2006. RFC 4271.

[30] 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.

[31] Telegeography. Bandwidth pricing report. `http://www. telegeography.com/product-info/pricingdb/ download/bpr-2009-10.pdf`. Retrieved: June 2011.