

Dynamic Capacity Contraction: A Framework for Congestion-Sensitive Pricing¹

Ranjeeta Sinha, Shivkumar Kalyanaraman, T. Ravichandran

Rensselaer Polytechnic Institute, Troy, NY 12180

E-mail: ranje@cs.rpi.edu, shivkuma@ecse.rpi.edu, ravit@rpi.edu

Abstract

This paper proposes a framework called "**Dynamic capacity contracting**" implementable using the differentiated services architecture in the Internet. A particularly interesting aspect of this framework is its use in implementing congestion-sensitive pricing. The central idea of congestion-sensitive pricing is a network that could, based upon congestion monitoring mechanisms, raise prices and vary contract terms dynamically. We develop a specific congestion-sensitive pricing scheme in the framework and present a preliminary performance evaluation of its technical and economic efficiency aspects, and position it relative to prior work by Clark and Varian /Mac-Kie Mason.

Key Words: Differentiated Services, Internet Pricing, Internet Economics, Congestion Control.

1. Introduction

Over the past ten years, the Internet has grown rapidly and diffused across a diverse user population [rai98]. Though technical measures for managing congestion are being developed at a rapid pace (eg: differentiated services [diffserv], TCP/IP improvements like SACK [tcp-sack], over-provisioned ISP cores etc), economic measures like responsive pricing have been proposed to achieve network and economic efficiencies [smart-markets, gupta97, expected-capacity].

Several schemes have been proposed to price the Internet and its various domains [smart-markets, expected-capacity, gupta97, kelly98]. However, there has been little experience in implementing and studying these schemes in the production Internet, though recent studies by Edell and Variaya [variaya99] are an important first step in that direction. A major impediment in this process has been the minimalist "best effort" service model of the IP, which does not provide a standard mechanism to specify packet forwarding behaviors other than the "best-effort" service. In other words, the lack of models for implementing pricing schemes using IP has meant that several of the proposed schemes have remained in the theoretical domain.

However, this scenario has recently changed as the Internet Engineering Task Force (IETF) has standardized two approaches to support service differentiation. The first approach is called Integrated Services ("int-serv") [intserv]. In this service framework, a signaling protocol called

¹ This work was funded by NSF Contract number ANI9819112, and an internal Rensselaer award

RSVP reserves network resources based on user defined service parameters. The second approach called Differentiated Services ("diff-serv") [diffserv] is expected to provide scalable service discrimination without the need for per-flow state signaling at every hop. While the two approaches can coexist and interoperate, it is expected that the latter approach (diff-serv) will be the choice of ISPs and backbone internetwork providers. In this paper, we focus on developing a pricing framework that utilizes the advanced traffic management features offered by the differentiated-services architecture.

The rest of the paper is structured as follows. First, we summarize the diff-serv architecture and highlight its traffic management features that are relevant to Internet pricing. Next, we outline our contribution: a *flexible framework for implementing a range of pricing schemes within the differentiated services architecture*. We go on describe a sample pricing scheme in this framework and its position relative to other pricing proposals in the literature, followed by some simple performance results to show the potential of this framework.

2. Differentiated Services Architecture and Control Mechanisms

The differentiated services model is shown in Figure 1: there are two types of routers, interior routers and edge routers. The interior routers need perform only a small set of highly optimized packet forwarding functions and classify packets based on the IP header field (known as "per hop behaviors" (PHBs)).

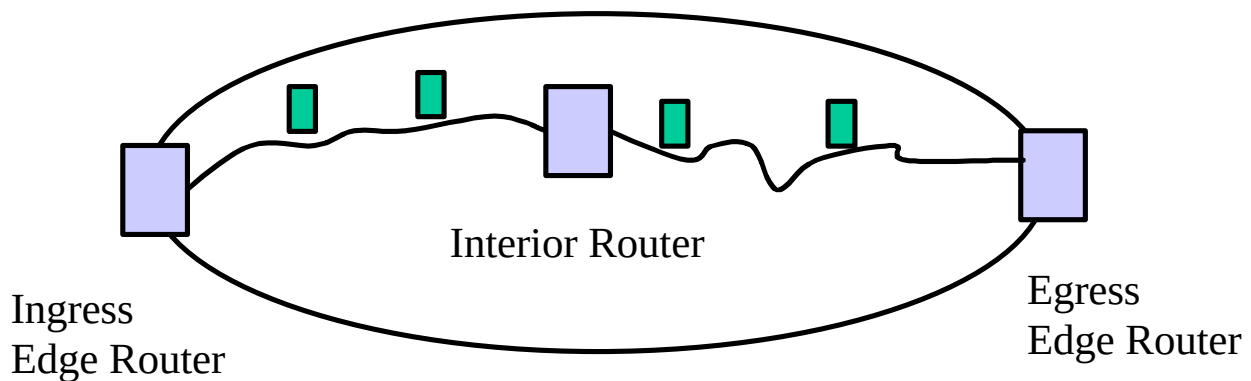


Figure 1: The Differentiated Services Architecture

Edge routers, however, participate in signaling, admission control, traffic conditioning and accounting functions including service level agreement (SLA) negotiation with customers. The

SLA is used as the basis for traffic conditioning (policing, marking, shaping) and accounting the actual traffic. Specifically, the ingress edge router marks the DS-byte in the IP header (formerly the IP TOS octet) of incoming packets indicating what per-hop behavior they should get at the interior routers. When these packets go through the interior routers, they get the per-hop behavior (PHB) as specified in the DS-byte.

Nichols, Jacobson and Zhang [bandwidth-brokers] proposed a control architecture that included a bandwidth broker for handling signaling and SLA negotiations between customers and providers. The bandwidth broker would be an software intermediary that is aware of the policies concerning the user and liaisons between the user and possibly multiple providers. COPS, a policy-based architecture is also being proposed for the control and provisioning of differentiated services networks [cops]. The differentiated services model combined with the proposed signaling/control schemes allows economic concepts such as "*price-based service differentiation*" and "*contracting*" to be incorporated within IPv4.

Kalyanaraman [kalyanaraman98] proposed a feedback mechanism to monitor congestion at the edges of the differentiated services network. In this mechanism an interior router, could mark a single bit in packets during congestion epochs, and the egress edge can use this to monitor congestion on edge-to-edge paths. In general, edge routers can collaborate to sense and exchange congestion information about paths, traffic aggregates or domains. Such information can then be used to apply controls on the flows experiencing congestion, or be incorporated into a pricing function. We explore the latter alternative in this paper, though work on the former alternative is also underway.

Based upon the above discussion, we note that the differentiated services model poses the following constraints on the deployment of pricing schemes:

First, diff-serv is built around the concept of aggregation, and hence it is important to formulate service properties which are not destroyed due to the operations of aggregation and de-aggregation.

Second, the asymmetry between edge routers and interior routers dictates that accounting and pricing mechanisms must be primarily concentrated near the edge routers with minimal participation of interior routers.

2. Related pricing proposals

Flat-rate price, the most common mode of payment today for bandwidth services, is popular for several reasons [flat-pricing]. First, flat rate pricing is efficient when the network is not congested, since most of the costs for network providers are sunk costs and unless there is congestion, the marginal cost of forwarding a packet is essentially zero [smart-markets, expected-capacity]. Second, flat fees are useful for prediction/forecasting for budgeting, require minimal accounting overhead, and encourage usage. However, flat rate pricing has problems. During congestion, the marginal cost of forwarding a packet is not zero, and flat pricing does not offer any (dis)incentive for users to adjust their demand, leading a potential ``tragedy of commons" [gupta97]

Two prominent proposals to deal with these problems are: 1) to regulate usage by imposing a fee based upon the amount of data actually sent (called **usage-based pricing**) and 2) use a fee based upon the current state of congestion in the network (called **congestion-sensitive pricing**). Usage-based pricing has several practical limitations. For example, usage costs are imposed regardless of whether the network is congested or not, and such schemes would have transaction costs. Moreover users don't like **a posteriori** pricing.

MacKie Mason and Varian [smart-markets] introduced the concept of ``congestion-sensitive" pricing in their scheme called ``**smart markets**." The actual price for each packet sent by an user is determined based upon the current state of network congestion. Users are expected to bid a price for their packets and packets whose bids exceeded some cutoff amount will be admitted and the rest are dropped or buffered. The cutoff amount is determined by the condition that the marginal willingness-to-pay for an additional packet is equal to the marginal congestion costs imposed by that packet. Congestion-sensitive pricing has also been used in roadways [ibd-1998]. Recently, MacKie-Mason, Murphy and Murphy [mmm97] also argued that closed-loop feedback about network congestion should be provided to users to allow them to adapt their demand.

Gupta et al [gupta97] proposed a **priority pricing scheme** where users' service requests are placed in priority queues, and an pricing strategy which sets the price for usage proportional to the average delay costs. Kelly et al [kelly98] propose a **proportionally fair pricing/rate-allocation scheme using shadow prices**. This scheme is motivated by the ATM ABR service and the use of shadow prices in telephony. Though it has strong fairness characteristics, is hard to deploy on the differentiated services model in the Internet.

Clark [expected-capacity] argued for direct ways of expressing desired network behavior (instead of bids and priority pricing). He proposed an “**expected capacity**” allocation scheme where users pay a price for a high probability of delivery for a given volume of traffic. If the actual usage exceeds the expected capacity during periods of congestion, users could experience a delay in the transmission. Clark et al's work also influenced the evolution of the differentiated services standard.

3. The Dynamic Capacity Contracting Framework

This section describes our framework qualitatively followed by the next section, which develops a sample scheme in this framework used in performance analysis. Our proposed framework is similar to the smart-markets proposal in the use of congestion-sensitive pricing, and the use of a bidding model for receiver-initiated services and uses feedback or exchange of congestion-information. The key differences are:

- a) The use of a **contract model** (similar to that suggested by Clark [expected-capacity]) instead of a bidding model. However we use **limited (or short-) term** contracts so that we can introduce a “**congestion-sensitive**” component into the contract.
- b) **Avoidance of usage or volume measurements**, and *a posteriori* pricing
- c) The use of a **congestion index** to measure the network state and mechanisms to assure better performance for users paying congestion-sensitive;
- d) The *granularity of contracting/accounting/pricing* is at a per-user or per-session level and not at a per-packet level.

We propose a contracting model like that of Clark [expected-capacity]. However, to implement congestion-sensitive prices, we need the flexibility to change the current price of a contract. Such change is not possible if all contracts are long-term, and therefore, we propose limited-term or short-term contracts, which by nature need to be dynamically renegotiated, and hence the name “**dynamic capacity contracting**”.

We assume for a given traffic class that the service negotiated in the contract is simply a function of volume (number of bytes), and the term of the contract (time units):

$$\text{Service (S)} = f(\text{Volume (V)}, \text{Term (T)}) \dots (1)$$

For shorter contract term-lengths, we simply assume that the user can burst upto the volume negotiated within the term of the contract. As in Clark's assured service model the provider will

assure that the negotiated traffic will be carried with a high expectation of delivery. In general, the user may send this traffic to any destination of its choice (i.e. a ***point-to-anywhere*** service); however for this paper, we focus on the case on ***point-to-point*** service since the measurement of congestion information in the latter case is non-trivial. We make one simplification to Eqn (1) by assuming that the term parameter (T) is fixed, i.e. different users cannot choose different term values. The user now sees a simple service offering: the flexibility to contract a desired volume (V) in the fixed term (T) at a given price per unit volume (Pv).

3.1 Congestion Index

Our model extends Clark's concept in that we will have a price per unit volume (Pv) based upon the service chosen (S), and a congestion-index (z), i.e.,

$$\text{Price per unit volume (Pv)} = g(\text{Service (S)}, \text{Congestion-index (z)}) \dots (2)$$

The congestion index (z) is a measure of congestion in the path (s) used by the customer's contracted traffic. Though we depict it as a single variable in the above equation, in general, it could be a set of variables. For example, measurements could be used to determine the fraction of the customer's contracted traffic sent on particular edge-to-edge paths, which could be used in turn to weight the per-path congestion index in order to find an overall congestion index. In other words, the congestion index can be customized on a per-customer basis. In this paper we use a simple per-path congestion index, and assume that each customer's contracted traffic is point-to-point to a single destination.

The use of the congestion-index (z) is inspired in part by work of Mac-kie Mason and Varian [smart-markets], Gupta et al [gupta97] and MacKie-Mason, Murphy and Murphy [mmm97]. Most of these authors calculate the equivalent of a congestion index per-packet and implement the clearing function (which matches the demand and supply based upon price or determines the price based upon demand and supply) at every bottleneck. Our model proposes that the congestion index (z) be used on a per-contract basis which means that it (z) can be determined without imposing participation requirements from interior routers, which may be bottlenecks. For this paper, we use a simple edge-to-edge feedback scheme, which involves some minimal interior node participation to estimate a congestion index. In future work we will eliminate the need for any

interior node participation in estimating z , and extend this work to cover “point-to-anywhere” type contracts.

The function g can also be dependent upon some estimate of demand elasticity which can be measured based upon the user’s choices of contracts. The demand elasticity is the slope of the curve between projected aggregate demand and price per unit volume. Note that Clark’s model can be thought of as a special case of our model when the term (T) is infinite, and there is no dependence upon the congestion-index (z). In other words, the price-per-unit volume is flat for the contract. Our model does not generalize to Varian’s model even if the contract term length is at a per-packet level because the congestion-index is collected at a time-scale of round-trip times. Also we use a contract model whereas theirs is a bidding model.

3.2 Demand curves and User Value Functions

When the user desires premium capacity, it approaches the provider to find out the current available prices for short-term contracts. The provider returns a table of available contracts and the price per unit volume (P_v) for each of them. The selection of contracts is then based upon user-perceived value functions (VF). All these transactions processed via the bandwidth-broker or policy control architectures [bandwidth-brokers, cops]. The value-function is a generalization of demand curves, including non-linear budget constraints, i.e., a function of user's long-term budget (B) and their desired average assured rate (R).

User value function (VF) = $h(\text{long-term budget } (B), \text{Desired Average Rate } (R))$

For example, a user would find little value if he/she has a high budget but can never obtain a desired average rate. This happens with flat-priced best-effort service. Another scenario is when the user's budget is exhausted and the user still gets far below its desired average assured rate. The latter scenario could happen because of price-fixing by a provider who has a monopoly, or if the user has under-budgeted. The monopoly problem could be alleviated by the use of bandwidth brokers which could access multiple providers leading to a more competitive environment (see Figure 2, next page). The study of user value functions is an open area which involves user behaviors, economic systems of competition/oligopoly/monopoly etc. We make some simplistic assumptions in this paper, but note that initial work in this area has been done by Edell and Variaya (INDEX project) [variaya99].

Observe that the bandwidth broker, a software intermediary, which is responsible for negotiating term contracts with providers on behalf of the users, can also maintain user costs to be

within a specified budget - a functionality that is similar to the Expenditure Controller Interface suggested by Danielsen & Weiss [danielson97]. We note that the providers can vary their advertised term prices based upon monitored congestion or time-of-day controls. A negotiated contract expires automatically after the term is over. This is consistent with the "soft state" approaches popular in several Internet protocols like DHCP and RSVP [rsvp]. In the future with Ipv6, the customer could have several alternative short-term contracts (and connectivity to multiple providers) in operation and on a per-packet basis choose a different provider [huitema98].

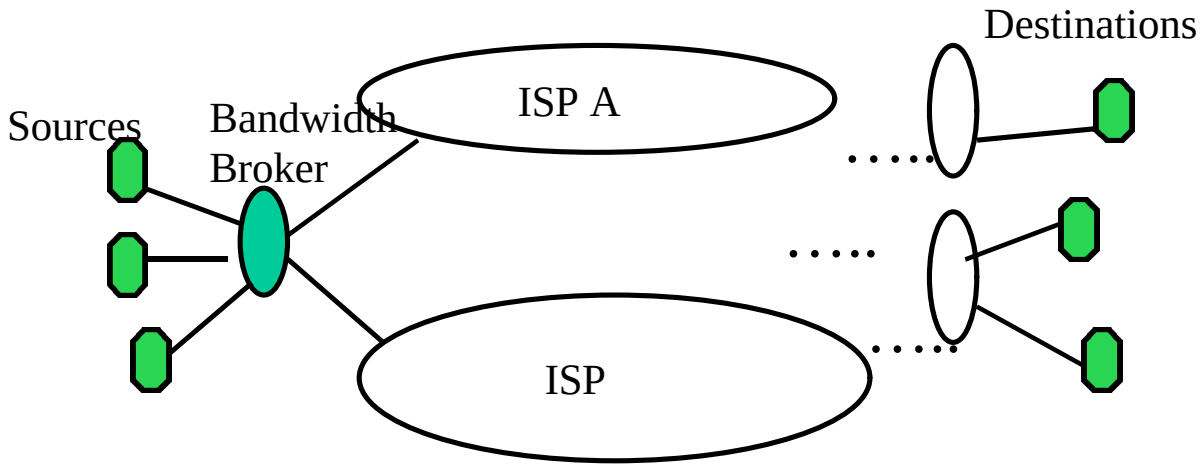


Figure 2: The Bandwidth Broker Concept

4. A Sample Scheme in the Framework

This section develops a simple scheme in the framework, which we also use in our performance evaluations. The development of schemes supporting a rich variety of contracts based upon this framework is a topic for future work.

In our scheme, the short-term contracts between a "customer" component and a "provider" component (at an edge router) are setup using a bandwidth broker. The customer initiates this process with a request for the table of short-term contracts available at the provider: a "send table" request. In response, the provider computes the entries of the table (price per unit volume, P_v) and returns the table to the customer. In this scheme, we assume a single entry in the table which specifies P_v for a contract of term length T . The price, P_v , is based upon a congestion-index (CI) as indicated in Figure 3. The congestion-index is a real-number between 0 and 1, and we assume a simple parabolic curve between P_{min} and P_{max} as congestion-index varies between 0 and 1. In

general, the curve could be based upon some estimate of demand elasticity on a per-customer basis. Observe that Dave Clark's expected capacity scheme [expected-capacity] is a special case of this curve, where the price-per-unit volume is constant relative to the congestion-index.

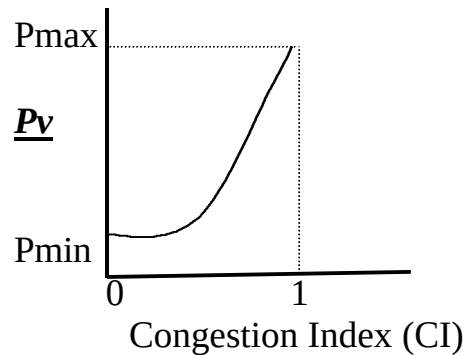


Figure 3: Price-per-unit-volume (P_v) vs Congestion Index (CI)

The customer chooses a desired volume of premium data traffic to be sent in time T based upon the price per unit volume, P_v , a demand curve, and its available budget. The demand curve is a curve between price and corresponding demand (volume): illustrated in Figure 4. We assume a simple exponential decay curve. An important economic measure is **consumer surplus**, which is the area under the demand curve, but above the current price value. Economic efficiency is achieved when consumer surplus is maximized.

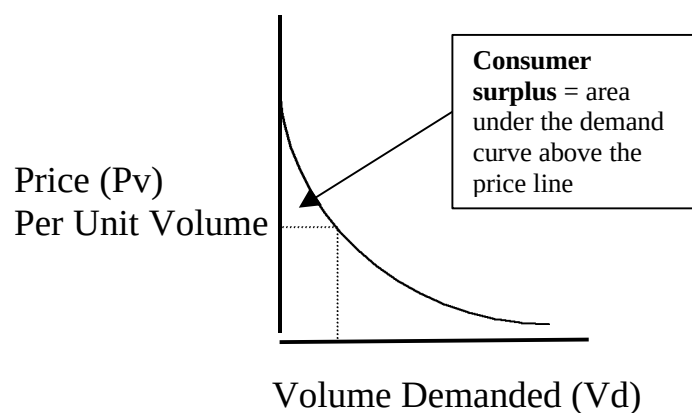


Figure 4: Demand Curve

Once a desired demand (V_d) is obtained from this demand curve, the customer checks if this can be paid for with the available budget, B . If not, it just chooses the amount permitted by the available budget:

```

If (Budget ( $B$ ) >
    (Volume demanded ( $V_d$ )* price per unit volume( $P_v$ )))
then
    Volume chosen ( $V$ ) = Volume demanded ( $V_d$ )
else if (Budget ( $B$ ) > 0)
Then
    Volume chosen ( $V$ ) = ( $B/P_v$ )
Else if (Budget ( $B$ ) == 0)
Then
    Volume chosen ( $V$ ) = 0
Endif
Decrease budget  $B$  appropriately

```

This choice of a volume is then conveyed to the provider, which sets up a leaky bucket (traffic conditioner) to mark upto V packets in time T as “IN” (high priority) and the rest as “OUT” (low priority). This step is the same as Clark’s model [expected-capacity].

We also add a simple **incentive scheme** to Clark’s model: OUT packets will be dropped at the edge itself with a probability = $(1 - CI)$. We immediately note that this incentive scheme is used to study only the degrees of freedom the framework can provide, since gratuitous loss-based incentives do not interact well with TCP flows. We are designing a better incentive scheme for TCP traffic which is based upon shaping the total output rate within bounds based upon congestion rather than dropping OUT packets. Therefore the performance results in this paper do not consider transport-layer interactions.

The packets thus conditioned enter the network and proceed through a series of interior routers in the differentiated services network till they reach the egress edge router (Figure 1). Similar to Clark’s model, we expect interior routers to provide support for service differentiation by using a **priority drop algorithm** proposed by Clark known as **RIO** (Random Early Drop, with IN/OUT marking) [expected-capacity]. RIO is a simple extension of the well-known RED drop algorithm [red], with different parameter sets for IN packets and OUT packets.

The final piece in the scheme is the calculation of the **congestion index (CI)** itself. Again, we propose a preliminary distributed algorithm which requires some simple support from interior routers and a bit (**explicit indication bit**) in packet headers. The bits being set aside for

differentiated services (the DS-byte [diffserv]) or ECN [ecn] could be used within the domain for this purpose. Future work will eliminate these requirements, and will also extend the concept beyond point-to-point contracts to include point-to-anywhere contracts. The interior routers participate by helping egress nodes identify the epochs when congestion occurs (called “**congestion epochs**”). The egress nodes have an **observation interval** (sized at the maximum edge-to-edge round trip time) and send a **bit-indication** to the corresponding ingress node only if congestion is detected in the observation interval. The ingress node maintains two intervals: one is the observation interval (same as the egress), and the other is a “**meta-observation**” **period** which spans multiple observation intervals (typically an order of magnitude more). *The congestion index (CI) is simply the fraction of observation intervals in the meta-observation period where a bit indication from the egress has been observed.*

The interior routers flag congestion epochs using a simple **ON/OFF marking technique**. When the queue size is above a given threshold, the explicit indication bit in **all** packets is set to 1; and when the queue size is below the threshold, all packets are passed through with the explicit indication bit unchanged. It is assumed that ingress routers initialize explicit indication bits to zero. The egress router detects congestion on the edge-to-edge path **once per observation interval** *if it observes any explicit indication bits set in packets belonging to the edge-to-edge aggregate flow*. Note that edge-to-edge aggregate flows can be identified through periodic exchanges between edge routers of information regarding unresolved flows (i.e. flows for which an edge-to-edge path is not known).

5. Performance Analysis

This section presents a preliminary evaluation of the scheme [expected-capacity]. Future work will also provide detailed modeling and quantitative comparison to Clark’s model and Varian/Mac-kie Mason’s smart-markets scheme. We use the following simple configuration in our analysis (Figure 5). It has a single bottleneck and two edge-to-edge aggregate flows corresponding to two customers. The bandwidth-broker (to negotiate and set up short-term contracts) is implemented between the customer and its connected ingress edge router. As described above, the interior router implements RIO, and the ON/OFF marker to support the calculation of the congestion index (CI). The destination nodes measure the **throughput** observed (ignoring transport

layer interactions) which is a measure of the “**technical efficiency**” of the scheme, i.e., its value as a traffic management technical tool. Measures of “**economic efficiency**” include the **consumer surplus** (described in section 4), the amount of **money made by the provider**, and the **transaction cost** (number of contracts/transactions made). Many of the scheme parameters (such as the curves used) are described in section 4. The duration of the short-term contract is equal to the length of the “meta-observation” period, and is set to 150 ms. The total simulation duration is 1 s which means that there are six (6) short-term contracts setup during the simulation. The observation intervals at the ingress and egress are set to 15 ms, which is the maximum edge-to-edge round-trip time. The bottleneck rate is 1 Mbps. Budget and money made by the provider are expressed in normalized units and bear a relation to real dollar amounts only in relative terms.

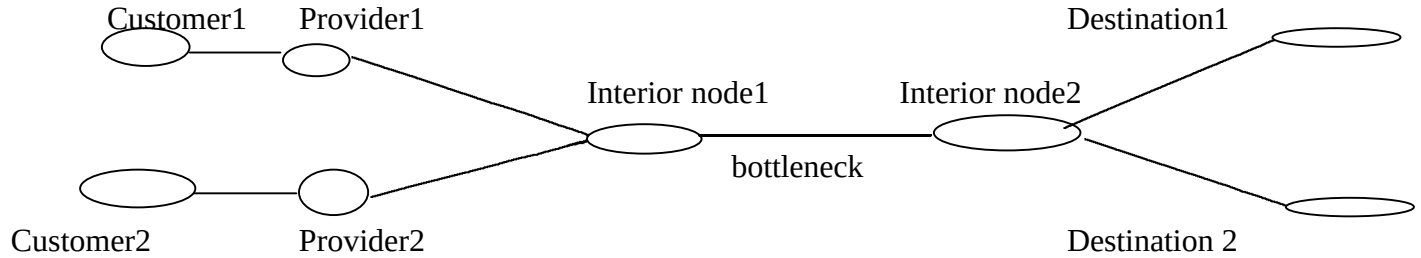


Figure 5: Single-bottleneck, 2-flows configuration

We run two experiments: the first where the two customers have an equal budget of 1000 units each, and the second where the first customer has a budget of 1000 units and the second customer has a budget of zero units. In the first experiment, we observe that:

- the congestion-sensitive pricing scheme is able to allocate throughputs proportional to prices paid, and manage congestion (**technical efficiency**),
- keep the consumer surplus (10000) and money made (1000 units) by the provider high, without incurring substantial transaction costs (6 transactions): an indicator of improved **economic efficiency**.

Experiment 1: Equal budgets of 1000 units each

Customers	Budget	Avg.Price Per unit volume	Transaction Count for each customer	Throughput Receiver 1	Throughput Receiver 2	Money made By each provider	Consumer surplus
Customer1 and 2	1000	218.80	6	483.2Kbps	492.8Kbps	1000	10,000

In the second experiment, we observe that throughputs are again allocated in proportion to the prices paid. Since the second customer had a zero budget, it does not get any throughput, and the entire bottleneck capacity is allocated to the first customer. Moreover, observe that the average price-per-unit-volume advertised to the second customer is higher. This is because it observes more congestion indications and hence computes a higher average price. If there is no feedback mechanism and congestion-sensitive pricing, we would not have seen this difference.

Experiment 2: Two customers: budgets of zero units and 1000 units respectively

Customer	Budget	Avg. Price per unit volume	Throughput at receiver	Money made by each provider	Consumer surplus
Customer1	1000	120	976.9Kbps	1000	10,000
Customer2	0	218.80	0Kbps	0	0

5. Summary

We have proposed a "dynamic capacity contracting" framework primarily inspired by the work of Clark [expected-capacity] and Mac-Kie Mason & Varian [smart-markets], and the differentiated services architecture [diffserv] which provides a platform for implementation. Distinguishing features of our work include the idea of "short-term" contracts, congestion-sensitive pricing, and a pragmatic focus on deployability. We have also proposed a sample scheme in this framework to illustrate the potential of the framework and show some sample performance results. We believe that such types of schemes could play a role in transitioning from today's completely flat-priced system towards a system that includes congestion-sensitive pricing for certain classes of service. Future work will include detailed comparisons with Clark's model and the smart market model, expansion of the concept of congestion-index to point-to-anywhere contracts, and refinement of the implementation to minimize participation by interior routers.

6. Acknowledgements

Our thanks to colleagues David Harrison and Murat Yuksel who assisted with the modeling, debugging and simulation aspects of this paper.

References

- [flat-pricing] Anania, L. A & Solomon, R. J (1997) Flat - The minimalist price, Internet Economics, Eds McKnight & Bailey, MIT press, 1997.
- [cops] Boyle, J., et al, "The COPS (Common Open Policy Service) Protocol," IETF Internet draft, Available from: <http://www.ietf.org/internet-drafts/draft-ietf-rap-cops-07.txt>
- [rsvp] Braden, R Zhang, L, Berson, S, Herzog, S & Jamin, S (1997) "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification," Internet Request For Comments (RFC) 2205, September 1997. Available from <http://www.cis.ohio-state.edu/htbin/rfc/rfc2205.html>
- [intserv] Braden, R, Clark, D & Shenker, S (1994) "Integrated Services in the Internet Architecture: an Overview," Internet Request For Comments (RFC) 1633, June 1994. Available from <http://www.cis.ohio-state.edu/htbin/rfc/rfc1633.html>
- [expected-capacity] Clark, D (1997) Internet cost allocation and pricing, in Internet Economics, Eds McKnight & Bailey, MIT press, 1997.
- [danielson97] Danielson, K & Weiss. M (1997) User control and IP allocation, Internet Economics, Eds McKnight & Bailey, MIT press, 1997.
- [ecn] Floyd, S (1994) "TCP and Explicit Congestion Notification" ACM Computer Communication Review, V. 24 N. 5, October 1994, p. 10-23. Available from <http://ee.lbl.gov/nrg-papers.html>
- [red] Floyd, S., and Jacobson, V. (1993) "Random Early Detection gateways for Congestion Avoidance". IEEE/ACM Transactions on Networking, V.1 N.4, August 1993, p. 397-413.
- [gupta97] Gupta, A, Stahl, D. O & Whinston, A, B (1997) Priority Pricing of Integrated Services Networks, Internet Economics, Eds McKnight & Bailey, MIT press, 1997.
- [huitema98] Huitema, c (1998) "IPv6: The new Internet protocol," Prentice-Hall, 1998.
- [ibd-1998] IBD (1998) "Freeways or Feeways", Investor Business Daily, 17th April 1998.
- [kalyanaraman98] Kalyanaraman et al (1998) "One-bit Feedback Enhanced Differentiated Services Architecture", draft-shivkuma-ecn-diffserv-00.txt>, March 1998 Available from <http://www.ecse.rpi.edu/Homepages/shivkuma>
- [kelly98] F. P. Kelly, A.K. Maulloo and D.K.H. Tan (1998) Rate control in communication networks: shadow prices, proportional fairness and stability, Journal of the Operational Research Society 49 (1998), 237-252.

[smart-markets] MacKie-Mason, J. K & Varian, H. R (1995) Pricing the Internet, in Public Access to the Internet, Kahin, Brian and Keller, James, ed., University of Michigan, Boston, Massachusetts, May 1993.

[mmm97] MacKie-Mason, Murphy, L & Murphy, L (1997) Responsive pricing in the Internet, in Internet Economics, Eds McKnight & Bailey, MIT press, 1997.

[tcp-sack] Mathis, M, Mahdavi, J Floyd, S & Romanow, A (1996) "TCP Selective Acknowledgement Options" Internet Request For Comments (RFC) 2018, October 1996. Available from <http://www.cis.ohio-state.edu/htbin/rfc/rfc2018.html>

[diffserv] S. Blake et al, (1998) "An Architecture for Differentiated Services," IETF Internet RFC 2475, December 1998. Available from: <ftp://ftp.isi.edu/in-notes/rfc2475.txt>

[bandwidth-brokers] Nichols, K, Jacobson, V and Zhang, L (1997) "A Two-bit Differentiated Services Architecture for the Internet", Internet Draft, <draft-nichols-diff-svc-arch-00.txt>, December 1997. Available from <http://www.ietf.org/internet-drafts/draft-nichols-diff-svc-arch-00.txt>

[rai98] Rai, A, Ravichandran, T & Sammadar (1998) Modeling the macro-diffusion of distributed IT infrastructures: The case of the Internet, Forthcoming, Communications of the ACM.

[variaya99] Richard J. Edell, Pravin P. Variaya, (1999), "Providing Internet Access: What we learn from the INDEX Trial," submitted to IEEE networks magazine. Available from: <https://www.INDEX.Berkeley.EDU/reports/99-010W/>