Achieving Optimal Revenues in Dynamically Priced Network Services with QoS Guarantees

Steven Shelford, Gholamali C. Shoja, Eric G. Manning
University of Victoria, Victoria BC, Canada
{sshelfor, gshoja, emanning}@cs.uvic.ca

Abstract

We have previously proposed the use of dynamically priced network services to provide QoS guarantees within a network. End-to-end QoS can be achieved by concatenating several of these services, perhaps from different ISPs. In this paper we consider the problem of a single ISP determining the bandwidth to allocate to each service, and on which path, in order to maximize revenue while guaranteeing end-to-end QoS. No knowledge of demand functions is assumed. Optimal allocation of bandwidth to services is first considered, where services are assumed to be routed on predetermined paths. We define the Iterative Allocation Adjustment heuristic, based on the concepts of tatonnement, which, through simulation, is shown to achieve over 95% of the optimal revenue for an ISP. We also examine how to value the links in the network to identify rerouting possibilities, or possible routes for new services, in order to improve the revenue of an ISP.

1. Introduction

Emerging broadband technologies promise faster download times and support for emerging streaming multimedia applications such as video-conferencing, telephony, and video-on-demand. Simple network over-provisioning is not enough because it assumes that provisioning will always outpace bandwidth requirements, and it ignores the reality that network bottlenecks can occur when switching is involved.

Internet applications requiring Quality of Service (QoS) suffer from the lack of QoS guarantees provided by the current Internet. This lack of guarantees limits both the performance of current applications and the emergence of multimedia and other application requiring QoS that can be developed and charged for.

In [1] we developed the concept of dynamically priced services, which we call QoS-Transit Services, to guarantee QoS. Each Internet Service Provider (ISP) offers to deliver traffic across its network only, with guaranteed performance, using QoS-Transit Services. Each service is defined by a particular source and destination pair on the network, along with a specified performance guarantee. Dynamic pricing is used to control traffic demand across each service and end-to-end QoS is achieved by sending traffic across several services, from different ISPs, in sequence. In [2] we illustrated how such end-to-end network QoS can be achieved across the Internet using existing protocol standards. Also, while pricing within a network may be dynamic, we note that such dynamics can be hidden from end users by each user's service provider.

The remainder of this paper is organized as follows. Section 2 discusses network interconnection agreements and bottlenecks. Section 3 defines QoS-Transit Services, the set of low-level services offered by an ISP to provide QoS guarantees. Section 4 defines the Iterative Allocation Adjustment heuristic which iteratively improves the allocation of bandwidth to the dynamically priced network services in order to maximize revenue for the ISP. This heuristic does not require knowledge of demand functions. Section 5 describes how to determine the value of links in order to identify rerouting possibilities, or to identify paths on which new services can be routed, in order to increase the ISP's revenue. Finally, section 6 concludes this paper.

2. Network interconnection

Most research in network pricing focuses on improving QoS for a single network [3-6]. To use these approaches on a set of interconnected networks would require much negotiation among ISPs, making the world of interconnection agreements even more complicated than it is now. Additionally, ISPs must be able to choose their own pricing schemes [7,8], as a means of market differentiation.

QShine'06 The Third International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks
August 7-9 2006, Waterloo, ON, Canada © 2006 ACM 1-59593-472-3/06/08…$5.00

Currently, ISPs connect through bilateral interconnection agreements of two general types: *peering* and *transit* [9-11]. A *peering agreement* between two ISPs allows each ISP to send traffic destined to a customer of the other ISP, directly to that ISP's network, for no charge. Peering is either *direct*, in which case a direct physical connection exists between the two networks, or via *exchange point peering*, where it occurs through a common exchange point, referred to here generically as an Internet Exchange.

A *transit agreement* between two network providers exists when one ISP provides another ISP with access to the entire Internet. The charging ISP accepts not only traffic for its internal customers, but also for traffic destined anywhere else on the Internet. Hence, the charging ISP must be able to deliver the traffic itself, or have agreements with other networks to deliver it.

We can classify ISPs as Tier-1, Tier-2, or Tier-3/Tier-4, as shown in Figure 1. There are no hard and fast rules on how to assign ISPs to these classifications. Tier-1 ISPs are often referred to as Internet Backbone Providers, as their networks make up the bulk of the Internet's core network. Tier-2 ISPs are often referred to as regional ISPs. Tier-3/Tier-4 ISPs are the remaining ISPs, which are smaller than regional ISPs.

Tier-1 ISPs do not enter into transit agreements to deliver any of their data. Tier-1 ISPs enjoy a mutual benefit by peering with other Tier-1 ISPs to extend their reach. This relationship is shown in Figure 1, where each of the Tier-1 ISPs peers with the others, forming a clique.

Tier-2 ISPs may also form peering relationships with one another when economically attractive. However, to reach the entire Internet for their customers, they must engage in a transit agreement with at least one Tier-1 ISP. Tier-1 ISPs have no obvious motivation to peer with a Tier-2 ISP, as the Tier-2 ISP would receive a disproportionate benefit from the relationship.

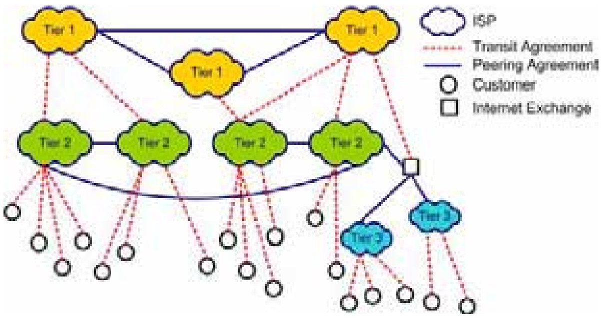

Figure 1. The hierarchical relationship of ISPs in the Internet.

Like a Tier-2 ISP, Tier-3 (and Tier-4) ISPs must connect with another ISP through a transit agreement. To reduce costs, Tier-3 ISPs can form peering relationships themselves, or aggregate their traffic to peer with a Tier-2 ISP. Such aggregation and peering is often done through an Internet Exchange, as illustrated in Figure 1.

Pan [12] and Akella [13] each investigated the locations of Internet bottlenecks. The studies revealed that Internet Exchanges and peering links are often the points of congestion. QoS-Transit Services can control the traffic across such bottlenecks, and by charging to use such links, or exchanges, ISPs may be motivated to properly provision these expensive resources when they are in high demand.

As for QoS, many ISPs providing transit services offer guarantees for latency, packet loss, and jitter. These guarantees, however, are with respect to performance across the ISP's own network and are generally stated as monthly averages only [14,15].

While it was traditionally understood that transit agreements must be entered into for a minimum of a month, the advent of *liquid bandwidth* is eroding this old-fashioned requirement. *Liquid bandwidth* refers to the ability to provision a connection within 48 hours. Provisioning times are dropping rapidly, thanks to the emergence of bandwidth market enabling systems such as Merkato [16].

Merkato is built upon the concept of a micro-market: a marketplace for a single commodity, such as bandwidth of a single link. Using this framework, QoS can be achieved by not overselling, and thus not over consuming, potentially scarce network resources such as bandwidth. Streaming Hand Services is an ISP that uses the Merkato platform to sell Tier-1 transit services with provisioning times as short as 5 minutes [17].

3. QoS-Transit Services

We define QoS-Transit Services as a set of primitive services offered by a network in order to guarantee the delivery of packets within the network. A QoS-Transit Service specifies a path across a network with stated performance criteria. We represent a QoS-Transit Service by a source, destination, maximum delay, maximum packet loss, maximum jitter, statistical calculation period, and a pricing function. We refer to an ISP that provides QoS-Transit Services as a QoS-ISP.

Figure 2 shows a network with 3 QoS-Transit

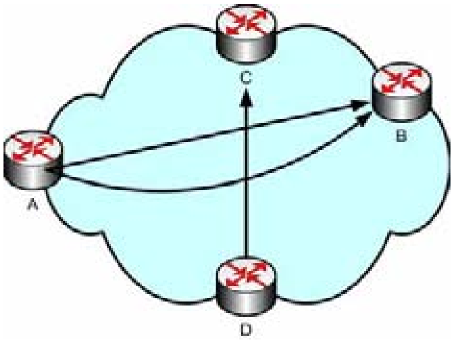

Figure 2. Sample network with 3 QoS-Transit services.

Services. The QoS-transit Services are defined between edge nodes and multiple services can exist between a pair of routers, as each service may guarantee different levels of QoS.

The pricing function defines the amount a customer is charged for a given amount of traffic given the current level of demand. The statistical calculation period defines the intervals over which time dependent values (such as data rates and other average values) are calculated. Currently, ISPs often calculate statistics using 5-minute samples. For dynamic pricing, we suggest much shorter sample sizes, such as 1 second.

To achieve end-to-end QoS, several QoS-Transit Services, each from a different ISP, may be concatenated. In [2] we described a feasible architecture which incorporates a scalable charging model and defines how routing between and across QoS-Transit Services can be achieved with existing standard protocols.

We carefully differentiate between *customers* and *end users*. *Customers* are entities that purchase QoS-Transit Services and *end users* are the individuals who are responsible for creating traffic. A customer purchases QoS-Transit Services for an aggregation of users. For example, a customer may be an enterprise or ISP, whereas the end user may be a single person within that organization or ISP. (While we do not disallow the possibility that a customer may be an individual end user, we expect that this will not be the usual case.)

QoS-Transit Services guarantee the delivery of packets within certain performance criteria. The QoS resulting from QoS-Transit Services does not guarantee bandwidth over the length of a session. Prices may increase, due to high demand, prompting a customer to cease the use of a service. However, recall that this behaviour is a low-level service, and that higher-level services such as admission control, multicast, and sender or receiver-pay billing can be implemented using these low-level services.

Higher-level services may be implemented either at the level of the user's ISP, or by other customers of QoS-Transit Services. A user's ISP may purchase the use of QoS-Transit Services in order to improve the QoS for certain users, or for certain traffic (i.e. multimedia traffic). As another example, a customer may be a multicast service provider, purchasing QoS-Transit Services to form multicast trees, and charging its users for use of those multicast trees. End users thus need not be aware of pricing fluctuations.

Throughout this paper we assume that the effective bandwidth of a service can be determined through measurement. Effective bandwidth can be understood as the bandwidth that must be reserved for the session in order for it to achieve its advertised performance guarantee. For example, a QoS-Transit Service with stringent performance guarantees may require a large amount of allocated bandwidth in order to achieve its stated performance in the event of traffic bursts. Additionally, the prices stated in the following sections are assumed to be normalized to a common price per unit of effective bandwidth (e.g. $/Mbps/minute).

4. Optimal allocation of QoS-Transit Services

This section describes an iterative process to arrive at the *optimal allocation* of bandwidth for each service, without knowing the underlying demand functions of the QoS-Transit Services. The *optimal allocation* is defined as the allocation of bandwidth to services such that, at the current posted prices, the revenue of the QoS-ISP is maximized. QoS-Transit Services are assumed to remain routed on predetermined paths.

4.1. Problem Definition

QoS-Transit Services guarantee performance by assuring that resources are not over subscribed:

$$\forall \ell \in L, \sum_{q \in Q} \alpha_q A_{q\ell} \le C_\ell \, , \qquad (1)$$

where L is the set of links in the network, Q is the set of QoS-Transit Services offered by the ISP, C_ℓ is the capacity of link ℓ, α_q is the effective bandwidth of service q, and $A_{q\ell}$ equals 1 if service q traverses link ℓ, and 0 if it does not. Each QoS-Transit Service is controlled by a pricing mechanism that assures that demand for a service does not exceed the bandwidth allocated to that service.

Optimal allocation of a QoS-ISP's network is the allocation of bandwidth within the network to the set of offered QoS-Transit Services, routed on predefined paths, so that the revenue of the QoS-ISP is maximized. Let $p_q(x_q)$ be the inverse demand function for QoS-Transit Service q, which defines the value per unit of bandwidth for the service when there is demand of x_q. Mathematically, the QoS-ISP needs to solve

$$\text{maximize} \quad \sum_{q \in Q} p_q(\alpha_q), \tag{2a}$$

$$\text{subject to} \quad \sum_{q \in Q} \alpha_q A_{q\ell} \le C_\ell, \forall \ell \in L. \tag{2b}$$

However, realize that this is the optimal objective, and that the inverse demand function is not known.

There exists an implicit relationship between bandwidth allocation and pricing, as illustrated in Figure 3. Each QoS-Transit Service is managed by a pricing mechanism that adjusts the service's price in order to manage the demand, and thus manage the QoS guarantees received by the traffic transmitted across the service. If the bandwidth allocated to a QoS-Transit Service is changed, then the pricing mechanism will adjust prices in order to match the demand to the bandwidth allocated to the service.

4.2. Iterative Allocation Adjustment heuristic

The Iterative Allocation Adjustment heuristic determines near-optimal bandwidth allocations for the offered QoS-Transit Services using a process based on the concept of *tatonnement*. *Tatonnement* is an iterative process where prices are adjusted until supply equals demand.

We assume that the *marginal revenue* of each service can be estimated. *Marginal revenue* is the increase in revenue due to the sale of an additional unit of a commodity. For example, if we can sell 5 widgets at \$8/widget, but we can alternatively sell 6 widgets at \$7/widget, then our marginal revenue for the 6th widget is:

\$7/widget × 6widgets – \$8/widget × 5widgets = \$2.

Marginal revenue could be estimated from the most recent price change, or it could be simply estimated as the current price per unit. Given a price vector $\overline{p}$ such that the resources of the network are not over-utilized, we can iteratively lower the price of a subset of the services and raise the price of another subset of the services, in order to increase the total revenue of the system. If the total revenue of the system cannot be raised through such price adjustments, then the system has an optimal allocation of resources. (Note that networks facing continuous changes in demand will not remain in a state of optimal allocation.)

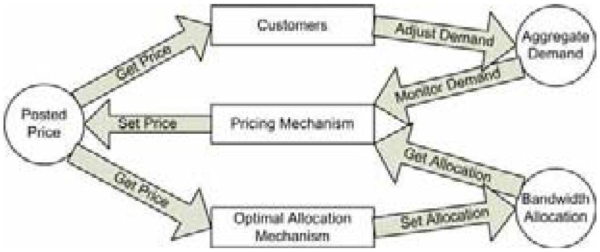

Figure 3. Relationship between the pricing mechanism and the optimal allocation mechanism.

Let v_q be the marginal revenue (the value) of service q. We can define the Iterative Allocation Adjustment heuristic precisely as:

1. Determine the marginal revenue for each of the services.
2. Maximize the marginal revenue gain obtained by adjusting each service's bandwidth allocation by at most b_q Mbps, where

   $$b_q = \begin{cases} b_{\max}, & w\alpha_q \ge b_{\max} \\ w\alpha_q, & w\alpha_q < b_{\max} \end{cases},$$

   and where w is the maximum percentage change per round for a QoS-Transit Service's allocation, up to at most $b_{\max}$ units of bandwidth. This can be solved as a linear program:

   $$\text{maximize} \quad \sum_{q \in Q} v_q x_q \tag{3a}$$

   subject to

   $$\sum_{q \in Q} A_{q\ell} x_q \le C_\ell - \sum_{q \in Q} A_{q\ell} \alpha_q, \forall \ell \in L, \tag{3b}$$

   $$-b_q \le x_q \le b_q, \forall q \in Q, \tag{3c}$$

   where the solved x_q is the change in bandwidth allocation to maximize revenue, given the constraints on maximum allocation changes per round.
3. Adjust the allocation for each QoS-Transit Service q, by x_q:

   $$\alpha_q \leftarrow \alpha_q + x_q, \forall q \in Q,$$

   where $-b_q \le x_q \le b_q$.
4. Allow the pricing mechanisms time to adjust the demands for the services to their new allocations.
5. Repeat.

The heuristic may be terminated if the revenue gain in a round, or over a number of rounds, is below some threshold.

Recall that the marginal revenues of the services are not known exactly, but are estimated. Also, the limits on allocation changes per round are to assure the stability of the system. Pricing mechanisms operate to assure that the demand for a QoS-Transit Service does not exceed its allocated bandwidth. Bandwidth allocation changes for a QoS-Transit Service must not exceed an amount that causes degradation to its guaranteed QoS.

Additionally, updating rounds must be staged far enough apart in time in order to assure the pricing mechanisms have time to adapt demand to the new bandwidth allocation.

4.3. Results

In this section we analyze the behavior of the Iterative Allocation Adjustment heuristic. The results were compiled for a 30-node network with 54 links where $b_{max} = 2\text{Mbps}$. QoS-Transit Services are randomly created and allocations are randomly assigned, though each service is assigned a minimum bandwidth of $C \cdot (2 \cdot |Q|)^{-1}$, where C denotes the capacity of each network link, and |Q| denotes the number of offered QoS-Transit Services.

Figure 4 shows how quickly optimality is reached for 50 services, with $w = 4\%$. 97.7% optimality was reached in 90% of test cases, as shown in Figure 5, and is reached after an average of 84 rounds. 90% optimality was reached by all test cases, after an average of 37 rounds, and 80% optimality was reached after only 13 rounds on average. These results are as expected, with earlier gains being achieved with less work than the subsequent gains.

Figure 6 shows the average gain in optimality per additional round depending on the optimality of a given solution. As we can see, the average gain diminishes as the solution approaches optimality. The implication here is that it may be difficult for a system to achieve a near-optimal (say greater than 95% optimal) solution. There may be insufficient time to execute the number of rounds necessary to achieve a desired optimality before prices change, due to changes in demand resulting from external factors, thus obviating the solution. The smaller the average gain in optimality per additional round, the less the system has to dynamically change its prices (due to ongoing changes in demand) in order to offset the changes made by the optimal allocation mechanism.

The time to calculate the allocation change for each round is negligible compared to the time required by the pricing mechanisms to react to these changes. While a round of the Iterative Allocation Adjustment heuristic can be performed in a few milliseconds, we estimate that the pricing mechanisms of the QoS-Transit Services should be allowed several seconds to adjust to the change. The actual time given to the pricing mechanisms to alter demand to meet the new allocations will depend on the specific pricing mechanisms implemented. Additionally, the time

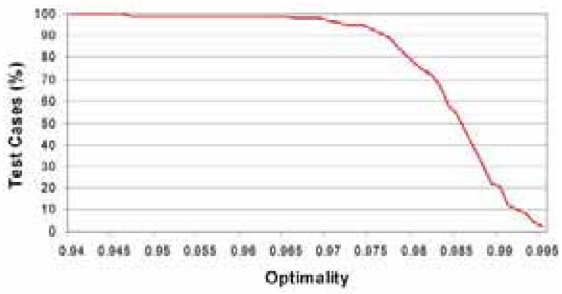

Figure 5. Percentage of test cases that achieve a specified optimality.

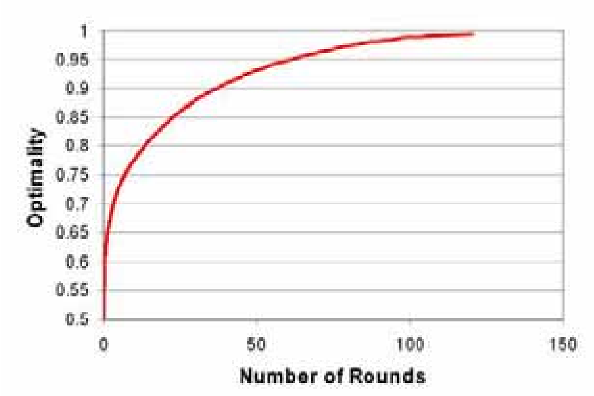

Figure 4. Optimality as a function of rounds for 50 services on a 30-node network, for *w* = 4%.

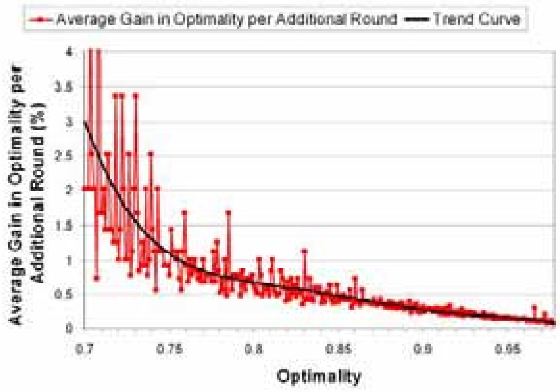

Figure 6. Average optimality gain per additional round for 50 services on a 30-node network, for *w* = 4%.

given to the pricing mechanisms will likely be a function of *w*: the more an allocation can change, the more time a pricing mechanism needs to adjust the demand to the new allocation. Thus, if each round occurs every 10 seconds, then 90% optimality will be reached in an average of 370 seconds. This is not exceptionally fast; however, we test from a random state, and the optimal allocation mechanism can be executed often to assure the allocations do not stray far from optimal.

While the previous results assumed 50 offered QoS-Transit Services, we now determine how the number of services offered by the system affects the number of rounds required to achieve a specified optimality. Figure 7 shows the number of rounds required to achieve 90% optimality for a varying number of QoS-Transit Services. As expected, the number of rounds required to reach a specified level of optimality increases with the number of QoS-Transit Services being offered. There are two reasons for these results. First, the fewer the services, the less likely it is that they will demand the same resources. The less they use the same resources, the closer to optimal their random allocation of bandwidth will be. Second, the more QoS-Transit Services, the less bandwidth each will be randomly assigned in our tests and the longer it will take for the higher priced services to be allocated a substantial proportion of the bandwidth.

Figure 8 shows the average number of rounds to achieve 90% optimality when *w* is varied. Not surprisingly, as *w* increases, the solution converges towards optimality more quickly. This is simply because the solution can change more in each round. However, the larger that *w* is, the more time a pricing mechanism may require to control demand to meet the allocation. Also, as *w* increases, so does the probability of violating QoS guarantees due to a large, sudden reallocation of bandwidth from a QoS-Transit Service.

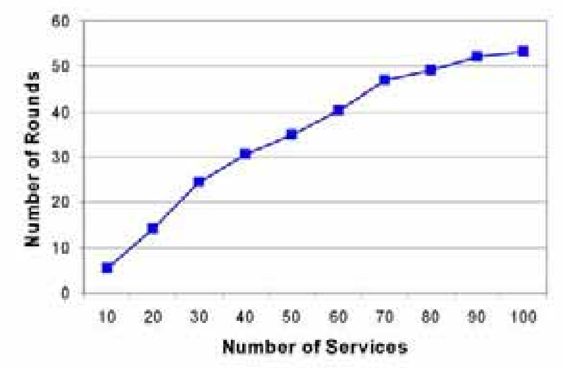

Figure 7. Average number of rounds to achieve 90% optimality with *w* = 4% on a 30-node network.

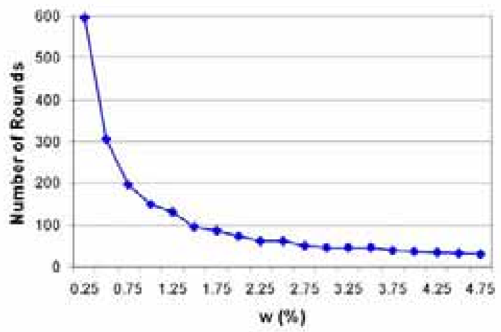

Figure 8. Average number of rounds to achieve 90% optimality for 50 services on a 30-node network.

5. Valuing links to identify reroute possibilities

While the QoS-Transit Services of a network may be optimally (or near optimally) allocated on a set of pre-defined paths, the paths on which they are routed across may not be optimal. In this section we identify how a network engineer can determine which links in the network are not carrying traffic with a high value. The engineer can then reroute certain QoS-Transit Services, increasing the overall value of the traffic transiting the network, and thus increasing the overall revenue for the QoS-ISP. We assume that the flow of a QoS-Transit Service is not split among multiple paths.

Given the set of QoS-Transit Services Q which are offered by a QoS-ISP, we can determine the value of the traffic that currently crosses each link within the QoS-ISP. Let the price, $p(q)$, for QoS-Transit Service q be normalized and reflect the price per unit of effective bandwidth. We can subsequently define a price to use a link to be $p(\ell)$, $\forall \ell \in L$, where L is the set of links with the network. As we will show, $p(q) = \sum_{\ell \in Lq} p(\ell)$ when the network bandwidth is optimally allocated, where L_q is the set of links that define the path that QoS-Transit Service q is currently routed on.

Consider two links (ℓ_1 and ℓ_2) of a QoS-ISP carrying 3 services (q_1, q_2, and q_3), as shown in Figure 9. $p(q_3) = p(q_1) + p(q_2)$ because we assumed

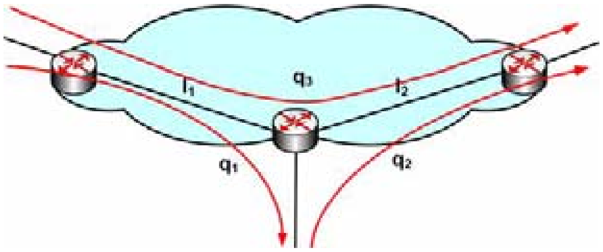

Figure 9. A network with 2 links carrying 3 QoS-Transit Services.

that the network bandwidth was optimally allocated. If $p(q_3) \neq p(q_1) + p(q_2)$, then an optimal allocation cannot exist because it would be more profitable to increase the bandwidth allocated to either q_3 or q_1 and q_2. Here, since q_1 and q_2 each traverse, and only traverse, links 1 and 2, respectively, we know that $p(l_1) = p(q_1)$ and $p(l_2) = p(q_2)$.

In the general case, we can define the price of each link in a network by solving the following set of linear equations:

$$\sum_{\ell \in Lq} p(\ell) = p(q), \forall q \in Q. \qquad (4)$$

We make two provisos in order to make the set of linear equations solvable under most conditions. First, if a set of links are connected in tandem and they carry the same QoS-Transit Services, then those links should be considered as one for the purpose of finding the price of the links. This proviso is illustrated in Figure 10.

The second proviso is the ignoring of links not used to capacity, assuming that only positive marginal revenues can occur. That is, the use of such a link is essentially given away free due to a lack of demand for it. For an example, consider the 5-node network with 4 QoS-Transit Services in Figure 11.

The set of linear equations resulting from Figure 11 is:

$$\begin{aligned} p(\ell_1) + p(\ell_2) &= p(q_1) \\ p(\ell_1) + p(\ell_3) &= p(q_2) \\ p(\ell_2) + p(\ell_4) &= p(q_3) \\ p(\ell_4) + p(\ell_5) &= p(q_4) \end{aligned}, \qquad (5)$$

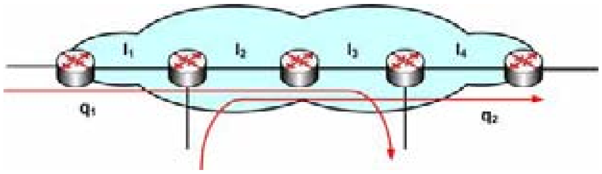

Figure 10. Links 2 and 3 here should be considered as a single link when finding the value of this network's links.

where $p(q_1)$, $p(q_2)$, $p(q_3)$, and $p(q_4)$ are known. If all the links in the network have the same bandwidth, and the bandwidth on links ℓ_3 and ℓ_5 are not fully consumed, then by applying the second proviso, we set $p(\ell_3)$ and $p(\ell_5)$ to 0. We then have:

$$\begin{aligned} p(\ell_1) + p(\ell_2) &= p(q_1) \\ p(\ell_1) &= p(q_2) \\ p(\ell_2) + p(\ell_4) &= p(q_3) \\ p(\ell_4) &= p(q_4) \end{aligned}, \qquad (6)$$

and thus,

$$\begin{aligned} p(\ell_1) &= p(q_2) \\ p(\ell_2) &= p(q_1) - p(q_2) \\ p(\ell_4) &= p(q_4) \\ p(\ell_4) &= p(q_3) - p(q_4) \end{aligned}. \qquad (7)$$

However, if the bandwidth of links ℓ_3 and ℓ_5 is fully consumed at the optimal bandwidth allocation, along with all the other links, then Equation 5 is unsolvable for this network. In these hopefully rare situations, we are unable to determine the value of the links.

By using the link valuation scheme a network engineer can determine which links are in high demand, and which are not. A network engineer can then route new services along less demanded paths, or perhaps reroute highly valued services along less demanded paths, in order to increase the overall profitability of the ISP.

A service on a path with a high valuation may be moved to a lower valued path with the expectation that the demand for that service will increase, and the ISPs revenue will increase. If demand increases for such services, then the value of the new path may increase as well, spreading the value of the QoS-Transit Services more evenly across the network.

There is too little information to determine an

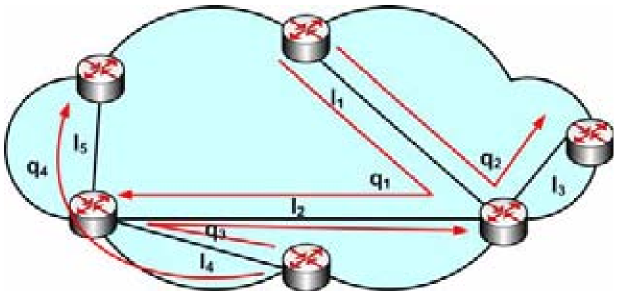

Figure 11. 5-link network with 3 QoS-Transit Services.

optimal strategy, unless the network provider can reasonably estimate the demand curves of the QoS-Transit Services. If so, the problem is a modified resource constrained optimization problem, as described and solved in [18].

6. Conclusion

If an ISP offers dynamically priced network services, such as QoS-Transit Services, to its customers, then it has an opportunity to maximize its profit by optimally allocating and routing those services. In this work, the ISP is not assumed to know, or be able to estimate, the demand functions for its offered services. We first proposed the Iterative Allocation Adjustment heuristic to find the optimal allocation of bandwidth for the offered services of an ISP, so that the ISP's revenue is maximized.

Using the Iterative Allocation Adjustment heuristic, we estimated that 90% optimal revenue, from a semi-random bandwidth allocation state, could be achieved in approximately 6 minutes on average. However, if the heuristic is executed often, without allowing the current allocation to deviate far from the optimal solution, then similar results will be achieved much faster.

To best route each of the services, we proposed that a network engineer determine the values of the links. The link valuations essentially equate to the prices of the links. By using the link valuation scheme, as proposed in this paper, a network engineer can determine which links are in high demand, and which are not. A network engineer can then route new services along less demanded routes, or perhaps reroute highly valued services along less demanded routes, in order to increase the overall profitability of the ISP.

References

[1] S. Shelford, E.G. Manning, and G.C. Shoja, "Achieving quality of service with overlay Internet service providers," *Technical Report DCS-300-IR*, University of Victoria, July 2005.

[2] S. Shelford, E.G. Manning, and G.C. Shoja, "A framework for quality of service control through pricing mechanisms," in *Proceeding of IEEE/IFIP NOMS 2006*, Vancouver, B.C., April 3-7 2006.

[3] J.K. MacKie-Mason and H. Varian, "Some economics of the Internet," Technical Report, University of Michigan, February 1994.

[4] A.M. Odlyzko, "A modest proposal for preventing Internet congestion," *DIMACS Technical Report 97-68*, AT&T Labs, October 1997.

[5] F.P. Kelly, "Notes on effective bandwidth," In Stochastic Networks: Theory and Applications (Editors F.P. Kelly, S. Zachary and I.B. Ziedins), *Royal Statistical Society Lecture Notes Series*, vol. 4, Oxford University Press, 1996, pages 141-168.

[6] M. Falkner, M. Devetsikiotis, and I. Lambadaris, "An overview of pricing concepts for broadband IP networks," *IEEE Communications Surveys*, vol. 3, no. 2, 2000.

[7] C. Courcoubetis and R. Weber, Pricing Communication Networks: Economics, Technology, and Modelling, England: John Wiley & Sons Ltd, 2003, pp. 21.

[8] S. Shenker, D. Clark, D. Estrin, and S. Horzog, "Pricing in computer networks: Reshaping the research agenda," *ACM Computer Communication Review*, vol. 26, no. 2, April 1996, pp. 19-43.

[9] P. Ferreira, "A model for interconnection of IP networks," Qualifier Paper, Carnegie Mellon University, January 6, 2003.

[10] C. Metz, "Interconnecting ISP networks," *IEEE Internet Computing*, vol. 5, no. 2, March/April 2001, pp. 74-80.

[11] W.B. Norton, "Internet service providers and peering," in *Proceedings of NANOG 19*, Albuquerque, New Mexico, June 2000,

[12] P. Pan, "Scalable resource reservation signaling in the Internet," Ph.D. Dissertation, Graduate School of Arts and Sciences, Columbia University, 2002.

[13] A. Akella, S. Seshan, and A. Shaikh, "An empirical evaluation of wide-area Internet bottlenecks," in *Proceedings of the 2003 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, San Diego, CA, USA, 2003, pp. 316-317.

[14] "Internet dedicated, global transit service level agreement," MCI, October 2004.

[15] "Service level agreement," infinity internet, online, May 13, 2005.

[16] "Merkato overview: a platform for real-time market-based network resource allocation," Technical Paper, InvisibleHand Networks, Inc., Revision 2, 2002.

[17] "StreamingHand: an in-depth look," Technical Paper, InvisibleHand Networks, Inc., Revision 2, May 23, 2002.

[18] S. Shelford, G.C. Shoja, E.G. Manning, "Optimal routing of dynamically priced network services," *Technical Report DCS-309-IR*, University of Victoria, February 2006.
