Two-Market Inter-domain Bandwidth Contracting

Abstract ID:1405

Abstract

An automated way of establishing inter-ISP contracts will enable ISPs to flexibly allocate their network resources to different contracts. However, this flexibility comes with the additional complexity in managing the ISP network and its risks. In this paper, we address how an ISP should assess the risk of a particular edge-to-edge path and reflect that risk on the contracting parameters, as well as respond to the risk by segmenting bandwidth between edge-to-edge short-term versus long-term contracts.

1 Introduction and Background

Recent killer applications such as IPTV, peer-to-peer file sharing and video on demand have started to reduce the bandwidth glut at the Internet backbone [1, 2, 3]. Though the Internet user community enjoyed a highly over-provisioned backbone in early 2000s, better bandwidth management technologies, such as multicast, are deemed necessary to handle the increasing demand arising from these recent applications [4]. Migration of data-rich applications such as IPTV provides more evidence that the trend of IP convergence will continue and the increasing demand for bandwidth is unlikely to completely vanish. In the backdrop of these trends, managing bandwidth allocation to customers presents a major challenge for network operators and received a lot of attention from the network researchers [5, 6] as well as practitioners [7] community.

Deriving more revenue by utilizing the available bandwidth is the essential focus of bandwidth management techniques. Many approaches have been explored involving advanced techniques, such as defining new types of contracts, like forward contracts, bailouts or options to achieve efficient contracting among ISPs. However, a crucial handicap of the existing inter-ISP economics is the coarse granularity of contracts (or service level agreements (SLAs)) among ISPs. SLAs take long time to establish or tear down, and are not dynamic enough to support healthy economics. The precursor to the realization of more dynamic and automated contracting is the sufficient motivation for ISPs to invest and install necessary tools and protocols to operate with such dynamic contracts. In this paper, we focus on answering the question of how much benefit there is if the contracts were classified in a simple two-market regime, where one market offers highly dynamic contracts with on-demand operation and the other offers long time-scale durable contracts.

One point of view is to have an ISP network participate in a set of complementary markets [8, 9, 10]. Mainly motivated by the fact that Internet bandwidth is not heavily utilized, the concept is to establish tools that will help emerge new complementary markets by using the leftover idle bandwidth. At the other end of the spectrum is the emergence of markets using heavily demanded parts of the network, but with explicit quality guarantees. Such complementary markets may have the potential to provide new values to the users, and thus offer profits that keep the ISPs motivated for further investment. A parallel issue to these advanced contracting techniques is the management of risks in ISPs' investments. Since the costs of bandwidth investments are enormous and typically "sunk", ISPs look for methodologies that can help them manage risks in such investments. Though monitoring and traffic matrix estimation techniques [11, 12] have been extensively engineered, inter-ISP contracting mechanisms aiming to help ISPs manage risks have only recently been explored [13].

In order to realize advanced contracting mechanisms, more temporal dynamism and spatial granularity in SLA establishment are needed. An automated and more dynamic way of establishing inter-ISP contracts will enable ISPs to flexibly allocate their network resources to different contracts. However, this flexibility comes at the cost of additional complexity in managing the ISP network. Several interesting questions arise in this context, such as, 1) How will an ISP decide when and where (i.e. between which ingress-egress pairs) to advertise a new contract? 2) How will an ISP assess the different risks involved and reflect them on the contracting parameters? 3) How will an ISP divide its links' bandwidth to different types of contracts? In this paper, we focus on a simplified version of the last question above, where only two types of contracts are considered, namely short-term contracts and long-term contracts.

Our approach is similar to the direction of [8], however, we strictly focus on how to divide the bandwidth available on an ISP network into the two segments, short-term and long-term, to maximize the aggregate revenue. To facilitate this analysis, we idealize the ISP network topology in terms of its physical, router-to-router, links, and utilize a certain set of links to create edge-to-edge contract paths. The decision variable is set for each contract path to determine the amount of bandwidth on that path will be set aside for the long- versus short-term demand. Utilizing a revenue generation capacity structure of short and long term contracts, aggregate revenue optimization formulation are developed and analyzed for the ISP's contracting strategy. The contributions of the paper are the following: We model and formulate the revenue maximization problem by considering the constraints imposed by the correlation among the contract links in an ISP network. Through a detailed simulation study, we provide several insights on the optimal reservation levels for the two types of contracts considered.

The rest of the paper is organized as follows: In Section 2.1, we describe the network model and the notations used. In Sections 2.2 and 2.3, we describe separately the problem formulations for the single and multiple edge-to-edge links, respectively. In Section 3, we present a detailed simulation-based study of the problem solution for different choices of the problem parameters. Finally, we provide concluding remarks and future directions in Section 4.

2 Problem Formulation

2.1 Network Model

We consider the contract-switched Internet architecture shown in Figure 1, in which, each ISP is abstracted as a set of edge-to-edge contract links, where and 'edge' refers to a node (ingress or egress) at which the ISP peers with some neighboring ISP. The edge-to-edge contract links are advertisable contracts between pairs of ingress-egress routers of an ISP. Let $\mathcal{I}$, $\mathcal{E}$ denote the set of ingress and egress routers of the ISP under consideration. Let $\{(i,e), i \in \mathcal{I}, e \in \mathcal{E}\}$ denote the set of contract links of the ISP. Each edge-to-edge contract link (i,e) is constructed using a set of physical (router-to-router) links of the ISP, and therefore represents a path between the ingress router i and the egress router e of the ISP. Let $\mathcal{L}$ denote the set of physical links of the ISP, and θ_{ie} denote the set of physical links that are used to construct the edge-to-edge contract link (i,e). We assume that the path θ_{ie} for each ingress-egress pair (i,e) is fixed and given. For example, path θ_{ie} can be the min-hop path between the ingress i and egress e, as shown in Figure 1. Thus, $\theta_{ie} \subset \mathcal{L}, \forall i, e$. Let C_l be the bandwidth capacity of the physical link l, where $l \in \mathcal{L}$. Each edge-to-edge contract link (i,e) also has a capacity, B_{ie}, that represents the total bandwidth units available for contracting along the contract link. Naturally, we define B_{ie} as the minimum of the capacities of the physical links on the path θ_{ie}, i.e., $B_{ie} = \min_{l \in \theta_{ie}} \{C_l\}$.

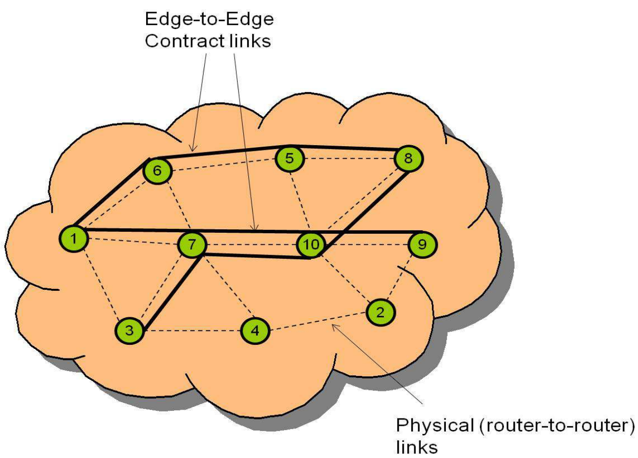

Figure 1: Contract-switching framework: An ISP is abstracted as a collection of several edge-to-edge contract links. The edge-to-edge contract links between ingress-egress pairs $(1,8)$, $(1,9)$, and $(3,8)$ are shown. The contract links $(1,9)$ and $(3,8)$ share the physical link $7-10$.

The ISP wishes to offer two types of contracts on each of its contract links, namely, short-term contract and long-term contract. To enable this, the ISP must segment the total available bandwidth on each of its contract links, B_{ie}, between the short-term and the long-term contracts. We assume that there is always enough demand for long-term contracts. The demand for short-term contracts, however, is assumed to be stochastic. The intuition behind this assumption is the expectation that long-term contracts are going to be made for stable traffic exchange cases where the contracting ISPs have a long-standing relationship and have a clear idea about how much traffic they will need to exchange. Such cases typically happens among ISPs at higher tiers of the Internet structure where traffic demand much more aggregated and stable in terms of direction and size.

Let the random variable Z_{ie} represent the demand for short-term contracts for the contract link (i,e). Let P_L denote the revenue from unit demand for long-term contracts per unit time. That is, a long-term contract will yield P_L units of revenue per unit time per unit bandwidth. Similarly, let P_S denote the revenue per unit bandwidth per unit time for short-term contracts. P_S would typically be higher than P_L, otherwise, the ISP has no motivation to offer short-term contracts. Further, short-term contracts will typically be associated with demand spikes arising from peak events such as Presidential speech or football match.

We initially consider a simplified version of the problem where there is a single edge-to-edge contract link in the ISP. Note that this will not be a typical case, since each ISP would often advertise several edge-to-edge contract links. However, when there are multiple links, each edge-to-edge contract link sometimes operate independently of the others, and therefore they can be studied individually. Therefore, we first present the problem formulation for single edge-to-edge contract link. In the next step, we consider multiple edge-to-edge links and extend the problem formulation to this setting, in order to study the impact of path dependency on the contracting strategy.

2.2 Single Edge-to-Edge Contract Link

Let there be a single edge-to-edge contract link in the ISP, with a total available capacity of B bandwidth units. The ISP wishes to offer short-term contracts and long-term contracts along this contract link. Let X be the amount of bandwidth reserved for long-term contracts. The remaining bandwidth on the edge-to-edge link, i.e $B-X$, would then be available for short-term contracts. We assume that there is always enough demand for long-term contracts, so that the total reserved units for long-term contracts, X, is used up. However, the remaining bandwidth $B-X$ may or may not be fully used depending on the amount of short-term demand Z. The ISP must choose an optimal value for X, such that the total expected revenue per unit time is maximized.
The total expected revenue per unit time is given by,

$$R(X) = P_L \times X + P_S \times E\left[min\left(B-X, Z\right)\right] \tag{1}$$

The revenue maximization problem can be stated as,

$$\max_X R(X)$$
$$\text{s.}t. \quad 0 \le X \le B$$

The solution to this problem depends on the value of P_L and P_S and also the characteristics of the short-term demand, Z, distribution. We solve the problem using MATLAB, by considering two probability distribution functions for the short-term demand, Z. The results are discussed in Section 3.1.

2.3 Multiple Edge-to-Edge Links

Next, assume that there several edge-to-edge links in the ISP's network. Let X_{ie} be the bandwidth reserved for long-term contracts along the edge-to-edge contract link, denoted by (i,e). When there are multiple edge-to-edge contract links, the paths represented by these edge-to-edge contract links often share physical (router-router) links between them. The physical links that are used by several edge-to-edge contract links impose a constraint on the long-term reservation levels (X_{ie}) of these edge-to-edge contract links.

Let $\vec{X} = \{X_{ie} : i \in \mathcal{I}, e \in \mathcal{E}\}$. Recall that θ_{ie} denotes the set of physical links that are used to construct the edge-to-edge contract link (i,e). Now, for each physical link $l \in \mathcal{L}$ (having a capacity of C_l), the set of edge-to-edge contract links whose paths use the physical link l is given by $\{(i,e) : l \in \theta_{ie}\}$. Therefore, the corresponding X_{ie}'s should satisfy the following constraint:

$$\sum_{\{(i,e):l\in\theta_{ie}\}} X_{ie} \le C_l \quad \forall l \in \mathcal{L} \tag{2}$$

The constraint states that the total capacity any link, l, can provide to all contract paths that use the link is at most the total link capacity, C_l. Each physical link l in the network introduces a linear constraint on the $X'_{ie}s$, therefore a total of $|\mathcal{L}|$ constraints are added to the problem.
The revenue from the long-term contracts per unit time is given by

$$R(\vec{X})^{long-term} = \sum_{i,e} P_L \times X_{ie} \tag{3}$$

It should be noted that the remaining bandwidth along the physical links (after reservation for long-term contracts) must be shared among the short-term demand of all the edge-to-edge contract links whose paths use these physical links. Therefore, the expected revenue from short-term contracts cannot be calculated using the straight forward approach described for single edge-to-edge link.

The expected revenue per unit time from short-term contracts, $R(\vec{X})^{short-term}$, given that X_{ie} units are already reserved for long-term contracts is obtained as follows. For each realization of the short-term demand, $\vec{Z} = \{Z_{ie} : i \in \mathcal{I}, e \in \mathcal{E}\}$, the short-term revenue is obtained by solving the linear optimization problem described below:

$$\max \quad \sum_{(i,e)} Y_{ie} \times P_S \tag{4}$$

$$\text{s.}t. \quad 0 \le Y_{ie} \le min(B_{ie} - X_{ie}, Z_{ie}) \quad \forall i, e \tag{5}$$

$$\sum_{\{(i,e):l \in \theta_{ie}\}} Y_{ie} \le \quad C_l - \sum_{\{(i,e):l \in \theta_{ie}\}} X_{ie} \quad \forall l \in \mathcal{L} \tag{6}$$

In above problem, Y_{ie} denotes the reservation for short-term contracts along the edge-to-edge contract link (i,e) for a particular realization of the short-term demand $\vec{z}$. Equation (5) states that each Y_{ie} should be less than the minimum of the demand Z_{ie} and the residual capacity of the contract link, $B_{ie} - X_{ie}$. In Equation (6), the term $(C_l - \sum_{\{(i,e):l \in \theta_{ie}\}} X_{ie})$ denotes the remaining capacity on the physical link l, that can be used for short-term contracts of all those contract links whose path uses the physical link l.

Let $R(\vec{X})^{short-term}_{\vec{z}} (= \sum_{(i,e)} Y^*_{ie} \times P_S)$ denote the optimal revenue from the above optimization problem (Equations 4 - 6). The expected revenue from short-term contracts, $R(\vec{X})^{short-term}$, is then $E_{\vec{Z}}[R(\vec{X})^{short-term}_{\vec{z}}]$, where the expectation is calculated with respect to the distribution of the short-term demands.

The ISP's revenue maximization problem can now be stated as follows:

$$\max_{\vec{X}} \quad R(\vec{X})^{long-term} + R(\vec{X})^{short-term} \tag{7}$$

$$s.\text{t.} \quad 0 \le X_{ie} \le B_{ie} \quad , \forall i, e \tag{8}$$

$$\sum_{(i,e):l \in \theta_{ie}} X_{ie} \le C_l \quad , \forall l \in \mathcal{L} \tag{9}$$

Note that the expected short-term revenue, $R(\vec{X})^{short-term}$, for each $\vec{X}$, is obtained by solving the optimization problem in Equations 4 - 6 and calculating the expectation.

Solving the revenue optimization problem (Equations (7)-(9)) for multiple edge-to-edge links is challenging, due to the presence of the expected short-term revenue term, $R(\vec{X})^{short-term}$. Unlike the single edge-to-edge link problem, the expected short-term revenue in this case can be obtained only solving a stochastic optimization problem, in which the constraints as well as the objective function are stochastic. Solving this problem analytically for a given short-term demand distribution is non-trivial. Hence, we resort to the following computational approach. We generate a large number of samples for the short-term demand, $\vec{Z}$, and the solve the optimization problem (Equations 4 - 6) for each realization of the short-term demand, $\vec{z}$, to obtain the short-term revenue for this particular $\vec{z}$. We then take the average of the short-term revenues for several such samples to be the expected short-term revenue. We use the MATLAB "fmincon" routine to solve the optimization problems.

3 Simulation Experiments

3.1 Single Edge-to-Edge Link

We first study the solution of the single link problem defined in Section 2.2. The capacity of the single contract link is set to 10. We consider two distributions for the short-term demand: 1) Uniform distribution between 0 and 10, and

2) Truncated Gaussian distribution in the interval $[0,10]$, with mean 5 and standard deviation 1. Figure 2 shows the optimal long-term reservation on the contract link, X^*, as P_L is increased from 0, while P_S is fixed at 10. We observe that X^* increases linearly in P_L when the short-term demand distribution is uniform, but the increase in X^* is non-linear if the distribution is Gaussian.

Recall that P_L and P_S denote the revenue received per unit time from unit bandwidth of long-term contracts and short-term contracts, respectively. Suppose P_L is much lower than P_S and if the entire capacity of the contract link is available for short-term contracts, then the total expected revenue generated from the short-term contracts would be higher than long-term contracts. Due to this, the optimal long-term reservations remain low when $P_L << P_S$. However, when P_L increases, higher revenue can be generated from the long-term contracts if more capacity is reserved for them. Therefore, X^* increases with P_L.

The non-linearity in X^*, observed in Figure 2 for Gaussian demand, can be explained as follows. For the truncated Gaussian distribution considered, the short-term demand distribution is concentrated around the mean level of demand, taken to be 5, with tails near 0 and 10. Therefore, when the capacity available for short-term contracts (i.e $B-X$) is lowered slightly from B $(=10)$, the expected short-term revenue does not change much, since the short-term demand can still almost always be supported the residual capacity. Moreover, additional guaranteed revenue can be obtained if some capacity is made available for long-term contracts $(i.e. X > 0)$. Therefore, X^* shows a jump as soon as P_L increases from zero. Beyond this, X^* increases at a slow rate with P_L. If X^* increases rapidly, the capacity available for short-term contracts would decrease and approach the mean of the demand distribution. As a result, there is significant reduction in the expected revenue from short-term contracts. This decrease can be compensated by the revenue from long-term contracts only if there is significant increase in P_L. Clearly, when $P_L = P_S = 10$, the entire capacity is reserved for long-term contracts. In this case, each unit of bandwidth reserved for long-term contracts would generate a revenue of 10, whereas each unit of bandwidth reserved for short-term contracts would generate less than 10 units of revenue.

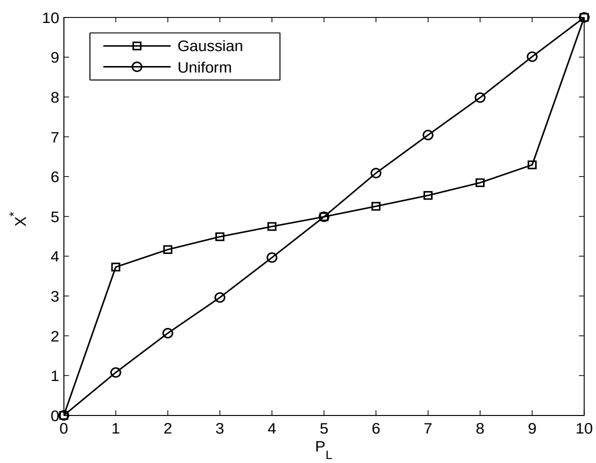

Figure 2: Optimal long-term reservation, X^*, on a single edge-to-edge link for different values of P_L. P_S is fixed at 10.

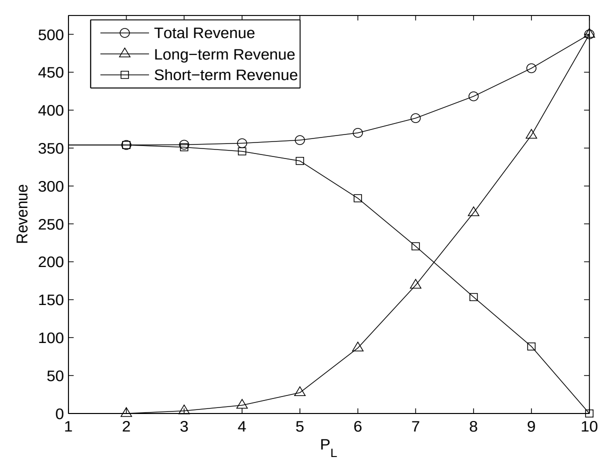

Figure 3: Figure showing the maximum total revenue, the maximum long-term revenue, and the maximum short-term revenue for a network setup with nine contract links for different choices of P_L. $P_S = 10$.

3.2 Multiple Edge-to-Edge Links

3.2.1 Network Setup

We consider the real topology map of the GEANT ISP [14]. The GEANT ISP topology consists of 23 routers. We choose three ingress and three egress routers randomly from the set of 23 available routers. We assume that an edge-to-edge contract link exists between each ingress-egress pair. Therefore, there are nine edge-to-edge contract links in the simulation setup. The capacity of each physical link is set to 10. Therefore, the capacity of each edge-to-edge contract link, B_{ie}, is also 10 based on our definition of B_{ie}. The short-term demand on each contract link is taken to be uniformly distributed between 0 and 10.

3.2.2 Results

The simulation results are shown in Figures 3 to 5. Figure 3 shows the maximum total revenue, the maximum long-term revenue, and the maximum expected short-term revenue as P_L is increased from 1. P_S is set to 10 as before. When $P_L \leq 5$, we observe the total revenue comes mainly from short-term contracts. This implies that the long-term reservation levels on all the nine contract links are close to zero for $P_L \leq 5$. This is simply due to the fact that P_L is not high enough to generate significant long-term revenue compared to short-term contracts. Moreover, since P_S is fixed, we also find that the total revenue (which comes mainly from short-term contracts) remains roughly at the same level for $P_L \leq 5$.

When $P_L > 5$, the long-term contracts generate higher revenue and therefore the long-term reservation levels along the contract links, X_{ie}, are also increased. The total revenue increases with P_L, since, long-term contracts are now the major source of revenue. The transition from purely short-term contracts to a mix of short-term and long-term contracts happens when $P_L = 5$. This can be explained as follows. Since $P_S = 10$ and the short-term demand distribution is uniform, each unit of short-term demand would generate an expected revenue of 5 ($P_S \times 0.5$). But, each unit of long-term demand would generate a revenue ($P_L \times 1$) higher 5, when P_L is greater than 5.

Figures 4 and 5 show the contract link capacity and the optimal long-term reservation levels on all the nine links for two different choices of P_L, while $P_S = 10$. From the two figures, we observe that increasing P_L from 5 to 10 increases the long-term reservation levels on the links as expected. However it is important to note that the long-term reservation levels on some of the contract links, namely $1,6,7$, in fact stays at zero for both the P_L values. Similarly, the long-term reservation levels on the contract links 5 and 8 are well below the maximum level of 10 units.

This feature of the solution is due to the constraint in Equation (9) introduced by the physical links that are shared among multiple edge-to-edge contract links. For the GEANT network setup considered, it is observed that each of nine contract links shared some physical links with at least one other contract link. For example, the edge-to-edge contract link 2 shared a physical link with the contract link 1, leading to the constraint $(X_1 + X_2 \leq 10)$. Moreover, the contract links 1, 4, and 7 also a shared physical link, leading to the constraint $(X_1 + X_4 + X_7 \leq 10)$. When we consider the solution shown in Figure 5, where $P_L = P_S = 10$, since $P_L = P_S$, the long-term reservations levels along all the links should be maximized in order to maximize the total revenue. So, if X_1^* was 10, X_2^*, X_4^*, X_7^* would have to zero to satisfy the two linear constraints. This solution leads to lesser revenue compared to solution $X_2^* = 10$ and $X_4^* = 10$, forcing X_1^*, and X_7^* to 0. Therefore, if a contract link appears in more of these feasibility constraints, a revenue maximizing solution will keep that contract link more for short-term contracts.

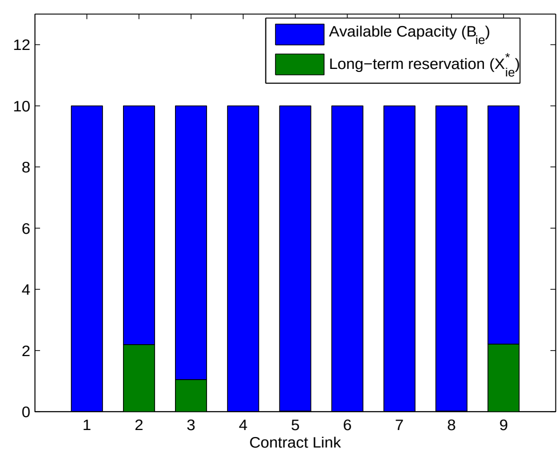

Figure 4: Optimal long-term reservation (X_{ie}^*) on each of the nine contract links when $P_L = 5$ and $P_S = 10$

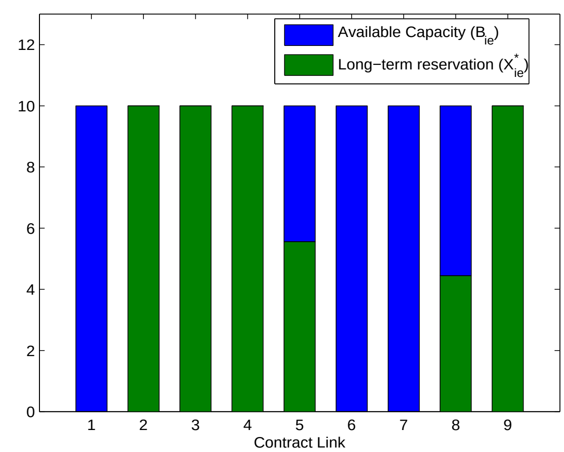

Figure 5: Optimal long-term reservation (X_{ie}^*) on each of the nine contract links when $P_L = 10$ and $P_S = 10$

4 Conclusion and Future Work

We formulated the revenue maximization problem for an ISP which wants to participate in two segments of bandwidth markets, that for short-term and long-term contracts. Our formulation aims to identify ISPs contracting strategy in the two markets and enables us to study the role of different parameters that affect the strategy. For instance, the level and distributional characteristics of short-term demand and the interactions among edge-to-edge contract links of the ISP's network are key determinants of the strategy and optimal revenue level. Through a detailed simulation study, we showed the optimal contracting levels and the revenues for several choices of the problem parameters. In our future exploration of this topic, we would consider the possibility of multiple paths underlying a given single contract link ingress-egress pair, instead of the single path assumption made in this paper.

References

[1] TeleGeography. The bandwidth glut is over. http://www.telegeography.com/cu/article.php?article_id=12155, April 2006.

[2] Stephen Lawson. Trans-atlantic internet cables may be filled by 2014. PC World, June 2009. http://www.pcworld.com/businesscenter/article/167143/transatlantic_internet_cables_may_be_filled_-by_2014.html.

[3] David Cotriss. Iptv and the looming bandwidth crisis. DailyIPTV, February 2008. http://www.dailyiptv.com/features/iptv-bandwidth-crisis-022808/.

[4] Gerard O'Driscoll. *Next Generation IPTV Services and Technologies*. Wiley-Interscience, 2008.

[5] C. M. Assi, Yinghua Ye, Sudhir Dixit, and M. A. Ali. Dynamic bandwidth allocation for quality-of-service over ethernet pons. *IEEE Journal on Selected Areas in Communications*, 21(9):1467–1477, 2003.

[6] Xipeng Xiao and L. M. Ni. Internet qos: a big picture. *IEEE Network*, 13(2):8–18, 1999.

[7] Michael J. Geile, Brian D. Anderson, Jeffrey Brede, Robert J. Kirscht, Michael J. Fort, and Mark D. Elpers. Dynamic bandwidth allocation. USPTO Patent 6279158, August 2001.

[8] MiklŮs Reiter and Richard Steinberg. Forward contracts for complementary segments of a communication network. In *Proceedings of IEEE INFOCOM*, pages 2705–2713, 2010.

[9] M. Yuksel, A. Gupta, and S. Kalyanaraman. Contract-switching paradigm for internet value flows and risk management. In *Proceedings of IEEE Global Internet Symposium*, 2008.

[10] E. Anderson, F. Kelly, and R. Steinberg. A contract and balancing mechanism for sharing capacity in a communication network. *Management Science*, 52(1):39–53, January 2006.

[11] Yin Zhang, Matthew Roughan, Carsten Lund, and David L. Donoho. Estimating point-to-point and point-to-multipoint traffic matrices: an information-theoretic approach. *IEEE/ACM Transactions on Networking*, 13(5):947–960, October 2005.

[12] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot. Traffic matrix estimation: Existing techniques and new directions. In *Proc. of ACM SIGCOMM*, 2002.

[13] W. Liu, H. T. Karaoglu, A. Gupta, M. Yuksel, and K. Kar. Edge-to-edge bailout forward contracts for single-domain internet services. In *Proceedings of IEEE International Workshop on Quality of Service (IWQoS)*, Enschede, Netherlands, June 2008.

[14] In *http://informatique.umons.ac.be/networks/igen/*.