Risk-Aware Network Profit Management in a Two-Tier Market *

Debasis Mitra[a] Qiong Wang[a]

[a]Bell Labs, Lucent Technologies, Murray Hill, NJ 07974,
mitra,qwang@research.bell-labs.com

We develop an optimization framework for the network service provider to manage profit in a two-tier market environment, a retail market where bandwidth is provisioned to serve uncertain demand and a wholesale market where bandwidth is bought and sold as commodity. Our model is built upon mean-risk analysis. We discuss the selection of a risk index that is consistent with stochastic efficiency criteria. We conduct numerical studies to investigate the influence of the network size and service provider's risk averseness on bandwidth management, and profit implications of the bandwidth wholesale market.

1. Introduction

This paper gives a framework for service providers' network profit management. The framework incorporates models of several facets of the task facing service providers. It also includes several key mechanisms for achieving management objectives. For instance, models and techniques for macro-level bandwidth management in support of these objectives are developed and presented here.

Uncertainty in (traffic) demand underlies the need for careful network profit management. In our models we use probabilistic distributions of demands as inputs. Obtaining such distributional information is in itself a major effort. On the other hand, forecasting and estimating traffic demand have been extensively studied, and much is known about relevant techniques [3], [5], [16].

Uncertainty breeds risk. A service provider must concern itself not only with mean profit and the strategies for its maximization, but also with the risk of profit falling below acceptable levels. Calculating risk and managing it are important components of network profit management, and also of this work. The optimization model in this paper incorporates a risk index. The rationale for its selection draws extensively from the mean-risk analysis that was originally developed in the finance community to address the needs of balancing growth and risk in portfolio management [6], [9]. As a recent derivative of the mean-risk analysis, robust engineering extends the approach to areas like power capacity expansion, aircraft scheduling, and minimum-weight structural design [11]. Risk in networking has its own characteristics. Risk is highly dependent on network scale, which is a theme that we develop further in our numerical studies. At the most basic level, the heterogeneity of demand between node pairs and capacity of links contributes to the variability of risk. Also, risk varies with service type, and, quite logically, this

*The authors acknowledge Alan Weiss for the benefit of discussions.

fact is reflected in market structure. This is exemplified by the two-tier market structure, which parallels a pair of contrasting services that is modeled here. Also note that service providers have varying attitudes towards risk, some being more averse to risk than others.

Considerable attention is given in this paper to find an appropriate measure of risk to use in the mean-risk approach. There are several candidates for such a risk index, which is incorporated in the objective function in the profit maximization. Some of these candidates, such as the variance of profit, lead to inferior solutions that are stochastically dominated by other feasible solutions. On the other hand, there are candidates, such as the Tail Value at Risk (TVaR), which are stochastically efficient but rather difficult to handle analytically and in the optimization. We show that the standard deviation of network profit is a good compromising measure of risk.

Turning to mechanisms for enabling risk-aware profit management, bandwidth management is at the top of our list. We note that in this paper, bandwidth sharing and management is at a high, macro-level, and rather different from that considered at the connection or packet levels. Stochastic traffic engineering [12] and the emerging technology of MultiProtocol Label Switching(MPLS) are primary enablers. The optimization in our model is with respect to the topology of the paths serving end-to-end demands and the amount of provisioned bandwidth on these paths, among other decision variables.

An important aspect of our model, and one that also leads to a mechanism for managing risk, is the two-tier market structure that was alluded to earlier. The first tier is a retail market where bandwidth is sold as a service. The second tier is a wholesale market where bandwidth is bought and sold as a commodity. The buying and selling of bandwidth is an important facility for bandwidth and profit management. Demand is deterministic in the wholesale market, so that there is no risk in profit management, but the price for carrying each unit of bandwidth is low. In the retail market the service provider can charge a premium price, but the demand is stochastic, so that there exists the risk of profit shortfall. By proper dimensioning of bandwidth provisioned for each market, the service provider can maximize mean profit at his/her acceptable risk level. We should point out that in our modeling framework the existence of the wholesale market, which is not widely established in today's marketplace, may be readily removed, in which case the stochastic demand in the retail market becomes the main focus. One of the goals of our work is to provide a platform for quantifying the cost-benefit tradeoffs that the wholesale market brings to profit management in telecommunications.

In a recent paper [12], we modeled a wholesale market in which the service provider is allowed to sell bandwidth. This paper extends the model by allowing the service provider to buy bandwidth as well. A central theme of [12] is the effect of demand uncertainty on network traffic engineering. In this paper the focus is different, with network profit management the dominant theme. Also new are the choice of risk index, the analysis of stochastic efficiency and the trade-off between the latter and tractability. The numerical studies of the implications of variations of service providers' risk averseness is new here. Also new are studies of the efficient frontier and the benefits of the wholesale market to service providers operating networks with diverse capacity levels.

The framework of this paper has the potential of providing the basis for studies of key future developments in the telecommunications industry. For instance, if the industry structure stabilizes to one with several competing service providers with distinct and possibly overlapping footprints, then a reasonable proposition is that, under certain conditions, resource sharing among all or a subset of these service providers will benefit not only the participating service providers but society and consumers also. Such resource sharing could take place under the aegis of multilateral agreements. Such agreements will also cover investments and profit sharing. Modeling frameworks will be needed to quantify the value proposition from such bandwidth sharing agreements. Also, effective techniques for allocating profits from use of shared resources will be needed [1], [4], [8]. We believe that the framework presented here can be extended and enhanced in natural ways to handle such challenges. In the present model the wholesale market with which the service provider interacts may be viewed as a proxy for the combination of all the other participating service providers. Clearly this view leads to dynamic, game-theoretic iterations composed of actions and reactions of the players, in which the basic moves by each player may be computed by an extension of the present model.

The paper is organized as follows. In Section 2, we present a general formulation of the problem. The modeling of the risk factor in the objective function is discussed in detail in Section 3. In Section 4, we develop several numerical examples to illustrate some implications of our model. Conclusions are given in Section 5.

2. Problem Formulation

2.1. Two-Tier Market

Our framework features a two-tier market structure. The first tier is a retail market where bandwidth is sold as services, such as voice, data, and video. Both the user base of the service provider and the usage rate of its customers are subject to random fluctuation, introducing uncertainties in the bandwidth demand and retail revenue. We reflect uncertainties in our framework by letting demands be random variables.

The second tier is a wholesale market where bandwidth is sold as standardized commodity, e.g., DS3 circuits. In our framework the market is an alternative source of revenue to the service provider where it can sell bandwidth wholesale. The market also provides a short-term supply of capacity to the service provider, where it can buy bandwidth to augment installed capacity in its network to serve retail demand. In this sense, the wholesale market provides a mechanism for many service providers to share resources and revenue, like the ones proposed in [1] and [8], and can be supported by an appropriate emerging engineering architecture [2].

We assume that the service provider under consideration is one of many participants in the wholesale market. Hence no service provider can dominate total demand and supply, and thereby significantly influence bandwidth price. This implies that the service provider's revenue from wholesale and cost of buying bandwidth follow deterministically from its buying and selling decisions.

As a result of guaranteed wholesale revenue, the service provider has no incentive to provision bandwidth to (uncertain) retail demand unless its unit revenue is higher than the wholesale price, which is the case we assume in this paper.

2.2. Admissible Route Sets

QoS and policy considerations are major constraints on provisioning decisions. The notion of *admissible route sets* allows these constraints to be taken into account in the optimization. Let $\mathcal{R}(v)$ denote the set of admissible routes for origin-destination pair v. Different admissible route sets for wholesale and retail services between the same node pair are allowed. For example, routes may be required to have lengths not exceeding specified thresholds, on account of propagation delay, and there may also be restrictions on the number of hops, since each hop is associated with an addition switching node and consequent incremental delay. The admissibility of a route may also depend on policy, which might reflect diverse considerations, such as security, the capability of switching nodes in the routes to handle certain services, link capacity, etc. Generating the admissible route sets is a substantial task in itself. In this paper, as in [10], we assume that these sets are given.

2.3. Model Formulation

In our formulation, the extent of the resources that the service provider can make available to itself is represented by the notion of *extended network*. An extended network is represented by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ and $\mathcal{L}$ are collections of nodes and links, respectively. Given the opportunity of buying bandwidth, the link set $\mathcal{L}$ is a union of $\mathcal{L}_U$ and $\mathcal{L}_V$, where the former is the set of installed links on which the service provider owns deployed capacity, and the latter is the set of virtual links on which the service provider has the option to buy capacity. The two sets are not necessarily mutually exclusive. Let C_l be the capacity that the service provider owns on link l, which is an input parameter, and b_l be the amount of capacity that the service provider buys on link l, which is a decision variable. Then $C_l = 0$ for $l \in \mathcal{L} - \mathcal{L}_U$, and $b_l = 0$ for $l \in \mathcal{L} - \mathcal{L}_V$. Assume the market price for buying capacity on link $l \in \mathcal{L}_V$ is p_l per unit of bandwidth. Then the service provider incurs a expense of $\sum_{l \in \mathcal{L}_V} p_l b_l$, for buying bandwidth, and the amount of bandwidth on each link l at the service provider's disposal is $C_l + b_l$.

Let $\mathcal{V} = \{(v_i, v_j) : v_i \in \mathcal{N}, v_j \in \mathcal{N}\}$ be the set of all node pairs. $\mathcal{V}_1 \subset \mathcal{V}$ is the collection of node pairs between which there are retail demands, and $\mathcal{V}_2 \subset \mathcal{V}$ is the collection of node pairs between which wholesale of bandwidth is feasible.

Retail demand between $v \in \mathcal{V}_1$ is characterized by a random variable, with its Probability Density Function (PDF) denoted by $f_v(x)$, and Cumulative Distribution Function (CDF) denoted by $F_v(x)$. Let $d_v (v \in \mathcal{V}_1)$ be the amount of capacity provisioned to serve retail demand between v. The provisioned quantity, d_v, can be routed on one or more admissible routes. Denote the admissible route set for $v \in \mathcal{V}_1$ by $\mathcal{R}_1(v)$ and let $\xi_r (r \in \mathcal{R}_1(v))$ be the amount of capacity provisioned on route r. Then $d_v = \sum_{r \in \mathcal{R}_1(v)} \xi_r$. We require that $0 < \bar{d}_v \leq d_v$ for all $v \in \mathcal{V}_1$, where $\bar{d}_v$ is defined as the minimum amount of bandwidth that must be provisioned for retail demand between v to satisfy the grade of service (GOS) that the service provider offers to its retail customers. Denote by T_v the random retail demand between node pair v. Then $t_v(d_v) = \min(T_v, d_v)$ is the amount of retail demand that is actually carried between v. Note that the revenue earned by the service provider is based on the carried demand. Let $m_v(d_v)$ and $s_v^2(d_v)$ be the mean and variance of t_v,

$$\begin{aligned} m_v(d_v) &= \int_0^{d_v} x f_v(x) dx + d_v \bar{F}_v(d_v) = \int_0^{d_v} \bar{F}_v(x) dx \\ s_v^2(d_v) &= \int_0^{d_v} x^2 f_v(x) dx + d_v^2 \bar{F}_v(d_v) - m_v^2(d_v) = 2\int_0^{d_v} x \bar{F}_v(x) dx - m_v^2(d_v), \quad (1) \end{aligned}$$

where $\bar{F}_v(x) \equiv 1 - F_v(x)$. Note that

$$\frac{\partial m_v(d_v)}{\partial d_v} = \bar{F}_v(d_v) \geq 0, \text{ and } \frac{\partial s_v^2(d_v)}{\partial d_v} = 2(d_v - m_v(d_v))\bar{F}_v(d_v) \geq 0. \tag{2}$$

Let π_v be the revenue earned for each unit of retail traffic carried between v. The total revenue derived from serving retail demand between v is a random variable $\pi_v t_v(d_v)$, for which the mean is $\pi_v m_v(d_v)$ and the variance is $\pi_v^2 s_v^2(d_v)$.

Similarly, let y_v be the amount of bandwidth provisioned for wholesale between $v \in \mathcal{V}_2$, $y_v = \sum_{r \in \mathcal{R}_2(v)} \phi_r$, where $\mathcal{R}_2(v)$ is the admissible route set for $v \in \mathcal{V}_2$ and ϕ_r is the amount of provisioned bandwidth on route r to carry wholesale traffic. Suppose e_v is the unit wholesale price between node pair v. Then the wholesale revenue is $e_v y_v$.

The service provider's profit equals combined revenue from serving retail demand and selling capacity, minus the cost of buying bandwidth,

$$W = \sum_{v \in \mathcal{V}_1} \pi_v t_v + \sum_{v \in \mathcal{V}_2} e_v y_v - \sum_{l \in \mathcal{L}_V} p_l b_l. \tag{3}$$

Due to random retail demand, $t_v (v \in \mathcal{V}_1)$ are random variables, and consequently so is W. We formulate the service provider's objective, denoted by Θ, as a function of W. The formulation of Θ should reflect service provider's desire to increase average profit and reduce the risk of profit shortfall due to the randomness of W.

To summarize, the above discussion leads to the following optimization model:

$$\max \Theta(W(d_v, y_v, \xi_r, \phi_r, b_l)) \tag{4}$$

subject to:

$$\sum_{r \in \mathcal{R}_1(v)} \xi_r = d_v \ (v \in \mathcal{V}_1), \ \sum_{r \in \mathcal{R}_2(v)} \phi_r = y_v \ (v \in \mathcal{V}_2), \tag{5}$$

$$\sum_{r \in \mathcal{R}_1(v): l \in r} \xi_r + \sum_{r \in \mathcal{R}_2(v): l \in r} \phi_r \leq C_l + b_l \ (l \in \mathcal{L}), \tag{6}$$

$$0 < \bar{d}_v \leq d_v \ (v \in \mathcal{V}_1), \quad 0 \leq y_v, \xi_r, \phi_r, b_l, \ (v \in \mathcal{V}_2, r \in \mathcal{R}_1(v) \text{ or } \mathcal{R}_2(v), \ l \in \mathcal{L}_V). \tag{7}$$

We defer the discussion of the objective function to the next section. As for constraints, (5) specifies bandwidth provisioning for both retail and wholesale demands on admissible routes; (6) limits the total amount of provisioned bandwidth from exceeding the sum of installed and purchased link capacities. In (7), $\bar{d}_v$ is defined as the minimum bandwidth that must be provisioned for retail demand between node pair v. The value of $\bar{d}_v$ is determined by the grade of service that the service provider offers to its retail customers. The equation also constrains all decision variables to be non-negative.

3. Modeling Risk

In the presence of demand uncertainty, maximizing the mean profit may not be the goal of the service provider. For example, a decision that guarantees a profit of \$99 may be more favorable than the one that generates profit of \$10000 with probability 0.01 and zero profit with probability 0.99, though the latter yields a higher mean profit. This means the risk of profit shortfall plays an important role in a service provider's decision-making, and should be accounted in the modeling. In this section, we develop an objective function that features both the maximization of mean profit and the containment of profit risk. We review the two relevant risk modeling frameworks in Section 3.1 and discuss our formulation in Section 3.2.

3.1. Relevant Frameworks for Risk Modeling

3.1.1. Mean-Risk Model

Mean-risk analysis, which has been widely applied to financial asset allocation, addresses the issue of risk averseness by offering a broader optimization objective. The approach starts with developing a risk index, which is a quantitative measure of the risk of profit shortfall, based on the distributional information. It then maximizes the weighted combination of the mean profit and the risk index, i.e., $mean - \delta * (risk\ index)$, where $\delta \geq 0$ is a parameter. Different levels of risk averseness can be reflected by choosing different values for δ. A higher value of δ indicates greater willingness to sacrifice the mean profit to avoid risk. On the other hand, by setting δ to 0, we can also include the case that the service provider maximizes the mean profit only.

3.1.2. Stochastic Dominance

Stochastic dominance theory defines a partial ordering of random variables based on their probability distributions [7]. Let W_1 and W_2 be two random variables, which, in our problem, represent profits under two different bandwidth management decisions. Then W_1 stochastically dominates W_2 to the first degree iff the former renders the service provider a better chance to exceed *any* profit target w, i.e.,

$$\forall w,\ Pr(W_1 \geq w) \geq Pr(W_2 \geq w). \tag{8}$$

Furthermore, W_1 stochastically dominates W_2 to the second degree iff

$$\int_w^\infty Pr(W_1 \geq \zeta)d\zeta \geq \int_w^\infty Pr(W_2 \geq \zeta)d\zeta \quad \forall w. \tag{9}$$

We consider a solution to our problem to be stochastically efficient if the corresponding profit distribution is not dominated in either degree. Note that it suffices to prove efficiency by showing that the profit distribution is not subject to second-degree dominance, which, by definition, is a necessary condition for dominance in the first degree.

3.2. Formulation of the Objective Function

We formulate the objective function as an instance of the mean-risk model with a constraint that the optimal solution is stochastically efficient. Whether the latter constraint can be satisfied depends on the characteristics of the profit distribution and the choice of the risk index. A common approach is to use variance as the risk index. If the profit is normally distributed, a solution that maximizes (mean - $\delta*$variance) is provably stochastically efficient [7]. For other distributions, the optimal solution to the mean-variance model can be stochastically dominated, which, for our problem, will be shown in Section 3.2.1. On the other hand, applying other risk measures that guarantee stochastic efficiency, such as Tail Value at Risk defined in Section 3.2.2, leads to models that are too complex to solve. As a compromise, in Section 3.2.3 we propose standard deviation as the risk measure.

3.2.1. Variance

Variance is a natural candidate for the risk index since it has been widely used in other mean-risk models in literature. Furthermore, using variance in our case leads to an objective function that is relatively easy to optimize. Nevertheless, as we will demonstrate in the following, the optimal solution to the mean-variance model can be stochastically dominated, meaning the solution is not stochastically efficient.

The objective function that uses the variance as the risk index can be written as

$$\Theta = E(W) - \delta Var(W) = \sum_{v \in \mathcal{V}_1} \pi_v [m_v(d_v) - \delta \pi_v s_v^2(d_v)] + \sum_{v \in \mathcal{V}_2} e_v y_v - \sum_{l \in L} p_l b_l. \tag{10}$$

Notice that (10) reflects the aforementioned assumption that demands between different node pairs are independent. Applying (1),

$$\frac{\partial \Theta}{\partial d_v} = \bar{F}_v(d_v)[1 - 2\delta\pi_v(d_v - m_v)] = \bar{F}_v(d_v)[1 - 2\delta\pi_v \int_0^{d_v} F_v(x)dx]. \tag{11}$$

The equation $2\delta\pi_v \int_0^{d_v} F_v(x) = 1$ has an unique solution, $\hat{d}_v$, such that

$$\frac{\partial \Theta}{\partial d_v} \geq 0, \ \frac{\partial^2 \Theta}{\partial d_v^2} \leq 0 \text{ if } d_v \leq \hat{d}_v \text{ and } \frac{\partial \Theta}{\partial d_v} < 0, \text{ if } d_v > \hat{d}_v. \tag{12}$$

Based on this observation, we impose $\max(\bar{d}_v, \hat{d}_v)$ as an upper bound on d_v, which makes our model a concave maximization problem that can be solved efficiently. If $\hat{d}_v \leq \bar{d}_v$, then $d_v = \bar{d}_v$ by (7). Otherwise, $d_v \leq \hat{d}_v$, which is a new constraint that reduces the original feasible region (a polyhedron defined by (5), (6), and (7)) to a convex set on which the objective function (10) is concave. This additional restriction on d_v does not affect the optimal solution since by (12), increasing d_v beyond $\hat{d}_v$ only reduces the value of Θ.

The above analysis implies that the optimal solution of the mean-variance model can be stochastically dominated. Since the optimal value of d_v is bounded by $\max(\bar{d}_v, \hat{d}_v)$, when the network has more than enough installed capacity to serve every retail demand to its upper-bound and wholesale is limited, some bandwidth will be left idle in the optimal solution. In this case, one can generate a new feasible solution by increasing the bandwidth provisioned to node pairs that are connected by underused links while keeping everything else unchanged. The new solution always has a better chance to get more profit than the optimal one and thus stochastically dominates the latter to the first degree.

3.2.2. Tail Value at Risk

Besides variance, other distributional parameters have also been proposed as candidates for the risk index. Of particular interests are those that guarantee stochastic efficiency of the optimal solution. One example is the Tail Value at Risk (TVaR), defined as:

$$TVaR(p) = \int_0^p q_W(\eta)d\eta, \tag{13}$$

where $q_W(p) = \inf\{w|Pr(W \leq w) \geq p\}$ is the p-quantile of profit W [7], [14].

We do not use TVaR as the risk index in our model because it cannot be formulated as an explicit function of decision variables, which makes optimization computationally difficult. Nevertheless, the consistency between the TVaR and stochastic efficiency helps us to justify the use of standard deviation as the risk index, which we will discuss next.

3.2.3. Standard Deviation as Risk Index

Our choice of risk index is standard deviation of profit, $s(W)$, which is computationally more tractable for optimization than TVaR. For networks with few node-pair demands, an optimal solution to the mean-standard-deviation model may still be stochastically dominated. However, the model is asymptotically compatible with the efficiency criteria as the number of node-pair demands increases. This property serves our purpose well since bandwidth transport networks usually have tens or even hundreds of nodes, and thus a large number of node pairs.

As in (3), the random component of profit W is a summation of many *independent* random retail revenues, $\pi_v t_v(d_v)$. From (2) the variance of each revenue, $\pi_v^2 s_v^2(d_v)$, monotonically increases with d_v, and thus has a lower bound $\pi_v^2 s_v^2(\bar{d}_v)$ from (7), and an upper bound $\pi_v^2 s_v^2(+\infty) = \pi_v^2 var(T_v)$. Therefore, except for some artificially-constructed extreme demand distributions, it is generally true that for every feasible solution

$$\lim_{|\mathcal{V}_1|\to+\infty} \frac{\pi_{v'}^2 s_{v'}^2(d_v)}{\sum_{v\in\mathcal{V}_1} \pi_v^2 s_v^2(d_v)} = 0, \ \forall v' \in \mathcal{V}_1.$$

This means that Lindeberg's condition is satisfied, so one can apply the Central Limit Theorem as follows [13],

$$\frac{W - E(W)}{s(W)} = \frac{\sum_{v\in\mathcal{V}_1} \pi_v[t_v(d_v) - m_v(d_v)]}{\sqrt{\sum_{v\in\mathcal{V}_1} \pi_v^2 s_v^2}} \to N(0,1), \text{ as } |\mathcal{V}_1| \to +\infty. \tag{14}$$

where $N(0,1)$ is standard normal distribution. It follows that if $|\mathcal{V}_1|$ is sufficiently large,

$$\begin{aligned} TVaR(p) &= \int_0^p q_W(\eta)d\eta = \int_{-\infty}^{q_W(p)} x dF_W(x) \\ &\approx \int_{-\infty}^{q_N(p)} [E(W) + s(W)\eta]dF_N(\eta) = p[E(W) - \frac{e^{-q_N^2(p)}}{p\sqrt{2\pi}} s(W)], \end{aligned} \tag{15}$$

where $q_N(p)$ is the p-quantile and $F_N(\eta)$ is the CDF of $N(0,1)$. Since $e^{-q_N^2(p)/2}/(p\sqrt{2\pi})$ changes monotonically from 0 to $+\infty$ as p varies from 1 to 0, $\delta = e^{-q_N^2(p)/2}/(p\sqrt{2\pi})$ has an unique inverse $p(\delta)$. By (15), the solution that maximizes $E(W) - \delta s(W)$ also maximizes $TVaR(p)$ for $p(\delta)$, and thus is stochastically efficient.

Using the standard deviation as the risk index, the objective function in (4) becomes

$$\max \Theta(W(d_v, y_v, \xi_r, \phi_r, b_l)) = \sum_{v\in\mathcal{V}_1} \pi_v m_v(d_v) + \sum_{v\in\mathcal{V}_2} e_v y_v - \sum_{l\in\mathcal{L}_V} p_l b_l - \delta \sqrt{\sum_{v\in\mathcal{V}_1} \pi_v^2 s_v^2(d_v)}. \tag{16}$$

Though more complicated than the aforementioned use of variance, the optimization of (16) is still tractable, as shown in [12].

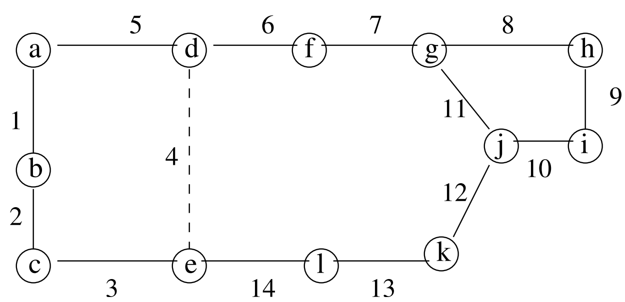

Figure 1. Network Topology

4. Numerical Studies

In this section, we discuss the implications of our model through numerical examples. We first describe the network topology and base case scenario in Section 4.1. In Section 4.2, we analyze the influence of the service provider's risk averseness on the optimization results based on the notion of efficient frontier. In Section 4.3, we examine the impact of the bandwidth wholesale market. In Section 4.4, we compare differences in optimal bandwidth management for networks of different sizes.

4.1. Framework and Base Case

We consider a sample network that has 12 nodes and 14 installed and virtual links. The network topology is shown in Figure 1, where link 4, represented by a dash line, is a "pure" virtual link that has no installed capacity. We assume that retail demands are symmetric in both directions and characterized by the Truncated Gaussian distribution, with support on the nonnegative real line and the following distribution function

$$Pr(T_v \leq x) \equiv F_v(x) = \frac{1}{\sqrt{2\pi}\sigma_v G_v} \int_0^x e^{-(\omega-\mu_v)^2/2\sigma_v^2} d\omega, \quad x \geq 0, \tag{17}$$

where the normalizing parameter is $G_v = Erfc(-\mu_v/(\sqrt{2}\sigma_v))/2$.

We assume $\sigma_v = \kappa\mu_v$ and let κ be the same for all v. For small value of κ (e.g., $\kappa \leq 1/3$) and positive μ_v, $G_v \approx 1$, in which case the truncation effect can be ignored and μ_v and σ_v are close approximations to the mean and standard deviation of the demand. In this case, κ is the coefficient of variation. We assume for all $v \in \mathcal{V}_1$, $\mu_v = \bar{\mu}$. Let h_v be the minimum number of hops between node pair $v \in \mathcal{V}_1$. Then $\sum_{v\in\mathcal{V}_1} \mu_v h_v = \bar{\mu} \sum_{v\in\mathcal{V}_1} h_v$. We define the ratio of this quantity to the total installed network capacity to be the network *load*, denoted by ρ, i.e., $\rho = (\bar{\mu} \sum_{v\in\mathcal{V}_1} h_v)/(\sum_{l\in\mathcal{L}} C_l)$.

We make the following assumptions to create a base case scenario.

1. All links has installed capacity 200 except link 4 which has zero installed capacity.
2. Retail demand exists between every node pair of the network. Bandwidth buying is allowed on every installed and virtual link, and wholesale is allowed between node pairs that are directly connected by these links. For demand distributions, $\rho = 0.7$ and $\kappa = 0.35$, which implies $\mu_v = 10.34$ and $\sigma_v = 3.62$ for all $v \in \mathcal{V}_1$.

3. The unit retail price $\pi_v = 50h_v$, where h_v is the minimum number of hops between node pair v. The unit wholesale price $e_v = 0.2\pi_v$. The unit price for buying bandwidth $b_l = 1.05e_{v(l)}$, where l is the direct link between v_l [2].
4. A path between node pair v is admissible for retail demand iff the number of links on this path does not exceed the minimum number of hops plus two, i.e., $h_v + 2$.
5. To satisfy GoS requirement, the amount of capacity provisioned to serve each retail demand should at least equal to the mean demand, i.e., $\bar{d}_v = E(T_v)$.
6. Unless otherwise noted, the risk parameter $\delta = 0.5$.

4.2. Efficient Frontier

Table 1 shows the influence of the service provider's risk averseness on key model outputs where δ ranges from a conservative value of 2.5 to the most aggressive value of 0 (in which case the service provider maximizes the expected profit only) in steps of 0.5. The service

Table 1
Changes of Optimal Decisions with Risk Parameter

δ	total buying expense	% wholesale bandwidth	wholesale revenue	expected retail revenue	% retail revenue
2.5	2656.4	24.7%	7034.3	84995.3	92.4%
2.0	2833.1	23.6%	6780.2	85638.5	92.7%
1.5	3020.9	22.7%	6566.1	86189.6	92.9%
1.0	3228.6	21.9%	6363.9	86703.8	93.2%
0.5	3428.0	21.1%	6169.4	87155.3	93.4%
0	3618.3	20.3%	5983.9	87548.7	93.6%

provider buys more bandwidth when it becomes less concerned about minimizing risk, as indicated by 36% jump (from 2656.4 to 3618.3) of the total buying expense in column 2. Allocation of total bandwidth to different markets is shown in column 3. In the most conservative case, 24.7% of total capacity is devoted to wholesale, which earns a lower but guaranteed unit revenue. This percentage is reduced to 20.3% in the most aggressive case. Consequently, the service provider's wholesale revenue drops 15% from 7034.3 to 5983.9 (column 4). This change makes the expected revenue from retail a more significant portion of total expected revenue, as shown in columns 5 and 6.

The efficient frontier in Figure 2 summarizes the cumulative profit impact of these changes. Each point represents an optimized combination of mean and standard deviation of profit at a given δ value. It represents the maximum expected profit (defined as reward) obtainable at a given level of risk (parameterized by the standard deviation), or the minimum risk the service provider has to take for a given reward.

[2]Letting the buying price to be slightly higher than the selling price avoids an unreasonable situation that it costs nothing for one to buy and sell bandwidth between the same node pair. The price difference can be interpreted as the transaction cost borne by the buyer.

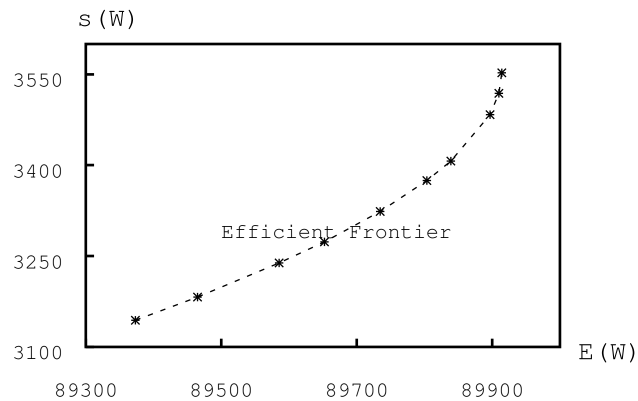

Figure 2. Efficient Frontier

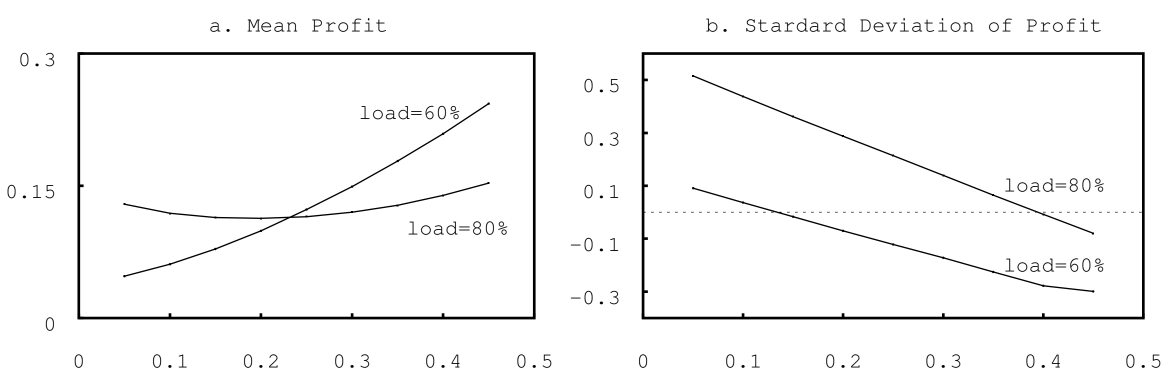

Figure 3. Percentage Change of Profit Distribution (x: e_v/π_v, y: $\Delta(E_W)$ or $\Delta(s_W)$)

4.3. The Value of the Wholesale Market

It is quite obvious that having a wholesale market is beneficial to the service provider. From the optimization point of view, the absence of the market is equivalent to restricting the general formulation by holding buying and selling quantities to zero. In the following, we discuss the influence of the price and traffic load on the benefit of wholesale.

We measure changes of mean and standard deviation of total profit by

$$\Delta(E_W) = \frac{E(\tilde{W}) - E(\bar{W})}{E(\bar{W})}, \ \Delta(s_W) = \frac{s(\tilde{W}) - s(\bar{W})}{s(\bar{W})},$$

where $\tilde{W}$ and $\bar{W}$ are optimized profit with and without the wholesale market, respectively. Figure 3a shows how $\Delta(E_W)$ varies with the wholesale price (given as a fraction of the retail revenue) under different network load conditions. When the price is low, the provider takes advantage of the wholesale market by buying cheap bandwidth to serve more lucrative retail demand. In this region, an increase in price reduces improvement of expected profit, as reflected by decreasing $\Delta(E_W)$. As the price keeps increasing, it eventually becomes profitable for the provider to sell more and buy less bandwidth in the wholesale market. At this point, $\Delta(E_W)$ starts to increase with the price. As the figure shows, the location of the switching point depends on the network load. In a lightly loaded network ($\rho = 0.6$), $\Delta(E_W)$ increases even at a very low price level (5% of unit revenue of retail demand). In a more heavily loaded network ($\rho = 0.8$), $\Delta(E_W)$ does not start to increase until the price reaches a relatively high level ($20\% - 30\%$ of retail revenue).

Figure 3b shows a decreasing trend of $\Delta(s_W)$ with the wholesale price. Low wholesale price results in a positive $\Delta(s_W)$, indicating profit variation increases as a result of buying bandwidth to serve more retail demand. As price increases, $\Delta(s_W)$ decreases and eventually becomes negative. This is a reflection of the fact that as bandwidth selling becomes more profitable, less capacity will be provisioned to retail demand, and variation in profit is reduced. Since the service provider that operates a more heavily-loaded network buys more and sells less capacity, $\Delta(s_W)$ is always smaller in a lightly-loaded network.

4.4. Influence of Network Scale

We now discuss differences in optimal bandwidth management for networks of various sizes. We start with the base case and scale installed capacity by a constant (γ) uniformly across all links. By keeping the load factor ρ unchanged, we scale the mean retail demand by the same factor. The scaling of the standard deviation is a little more complex. As the average volume grows, the variation of retail demands may or may not decrease. Similar to the "power law" in [3], we scale the standard deviation by γ^h, where $h \geq 0$. It is easy to verify that when $h = 1$, optimal values of all decision variables and the objective function change proportionately to γ. In the following example, we set $h = 0.5$ to model the situation of decreasing uncertainty to network scale.

We consider six differently sized networks ($\gamma = 0.5, 0.75, 1, 1.5, 2, 3$) and vary wholesale price as different fractions of retail revenue. We compare the provider's buying and selling activities in Table 2, where the normalized balance of wholesale bandwidth, defined as $(\sum_{v \in \mathcal{V}_2} y_v - \sum_{l \in \mathcal{L}_V} b_l) / \sum_{l \in \mathcal{L}_U} C_l$, is reported.

When the wholesale price is low, smaller service providers have negative balances, indicating that they are net buyers in the wholesale market. Larger service providers are net sellers who have positive balances. At the low price level, it does not cost much to buy enough bandwidth to accommodate the most optimistic demand scenario. Smaller providers buy even more because higher uncertainties in their retail demand imply a better chance of a large upward demand swing. As the price increases, smaller providers continue to buy more and sell less bandwidth than large providers, but the difference diminishes. When the wholesale price is high, providers of all sizes are better-off to provision capacity to retail demands to the level that is just enough to satisfy GoS, and wholesale the rest. In our example, the minimum quantity to satisfy GoS is set at mean demand, which is proportional to the network size. The assumption makes the proportion of bandwidth provisioned to retail demand more or less equal across networks of different sizes, which implies that their normalized wholesale balances become uniform. Similar to the previous subsection, we now consider benefits of wholesale market. The difference in buying and

Table 2
Normalized Wholesale Bandwidth Balance

	Network Scale					
e_v/π_v	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 1.0$	$\gamma = 1.5$	$\gamma = 2.0$	$\gamma = 3.0$
0.10	-12.13%	-4.18%	0.45%	5.80%	9.05%	12.89%
0.30	13.92%	17.12%	18.93%	20.90%	22.12%	23.57%
0.50	28.16%	29.47%	29.83%	29.98%	30.00%	30.00%

selling activities results in non-uniform $\Delta(E_W)$ and $\Delta(s_W)$ for differently sized networks. Being relatively heavy buyers, smaller providers enjoy larger improvements in mean profit when the retail price is low, as shown in Table 3. Larger providers are relatively heavy sellers who derive more significant benefits from the reduction of profit risk, which is shown by changes of standard deviation in Table 4.

Table 3
Changes $\Delta(E_W)$ for Different Size Networks

	Network Scale					
e_v/π_v	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 1.0$	$\gamma = 1.5$	$\gamma = 2.0$	$\gamma = 3.0$
0.10	11.56%	9.20%	7.95%	6.59%	5.85%	5.04%
0.30	11.35%	10.87%	10.64%	10.43%	10.33%	10.27%
0.50	18.05%	18.11%	18.10%	18.06%	18.06%	18.10%

Table 4
Changes of $\Delta(s_W)$ for Different Size Networks

	Network Scale					
e_v/π_v	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 1.0$	$\gamma = 1.5$	$\gamma = 2.0$	$\gamma = 3.0$
0.10	42.67%	34.99%	31.28%	27.32%	24.72%	21.21%
0.30	11.92%	7.63%	5.32%	2.64%	0.61%	-2.21%
0.50	-11.03%	-14.13%	-15.68%	-17.76%	-19.34%	-21.60%

5. Conclusion

We have developed an optimization model to support bandwidth management decision-making in a two-tier market environment. Our model is based on the mean-risk framework, and the stochastic efficiency of the optimal solution is strongly influenced by the choice of the risk index. We have discussed the tradeoffs of employing different distributional information to characterize risk and proposed the use of standard deviation of total profit. In our numerical studies, we have analyzed the impacts of service provider's risk averseness on various aspects of bandwidth management decision-making. We also have discussed profit improvement brought about by the presence of the wholesale market under various market and network conditions.

Our model can be extended and applied in several directions. It can be used not only to support decision-making of a single network service provider, but also to characterize interactions of multiple service providers in a wholesale market.

REFERENCES

1. E. Anderson, F. Kelly, and R. Steinberg, A Contract and Balancing Mechanism for Sharing Capacity in a Communication Network, *preprint*, 2002.
2. S. Biswas, D. Saha, and N. Mandal, Intercarrier Bandwidth Exchange: An Engineering Framework, *IEEE Communications Magazine*, vol. 41, no. 1, pp. 130-138, 2003.
3. J. Cao, D. Davis, S. V. Wiel, and B. Yu, Time-Varying Network Tomography, *Journal of the American Statistical Association*, **95**, pp. 1063-1075.
4. R. Gibbens, F. Kelly, G. A. Cope, and M. J. Whitehead, Coalitions in the International Network, *Teletraffic and Datatraffic - ITC13*, eds. A. Jensen and V.B. Iversen, Elsevier, Amsterdam, pp. 93-98, 1991.
5. M. Grossglauser and J. Rexford, Passive Traffic Measurement for IP Operations *INFORMS Telecom Meeting 2002*, Ft. Lauderdale, FL, March, 2002.
6. J. Hirshleifer and J. G. Riley, *The Analytics of Uncertainty and Information*, Cambridge University Press, Cambridge, England, 1992.
7. H. Levy, *Stochastic Dominance: Investment Decision Making Under Uncertainty*, Kluwer Academic Publishers, MA, 1998.
8. P. B. Linhart, K. G. Ramakrishnan, R. Radner, and R. Steinberg, Allocation of Value for Jointly Provided Services, *Telecommunications Sytems*, vol. 41, pp 151-175, 1995.
9. H. Markowitz, *Mean Variance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell, New York, 1987.
10. D. Mitra and K. G. Ramakrishnan, A Case Study of Multiservice Multipriority Traffic Engineering Design for Data Networks, *Proceedings of IEEE GLOBECOM 99*, pp. 1077-1083, December 1999.
11. J. M. Mulvey, R. J. Vanderbei, and S. A. Zenios, Robust Optimization of Large-Scale Systems, *Operations Research*, vol. 43, No. 2, pp. 264-281, March-April 1995.
12. D. Mitra and Q. Wang, Stochastic Traffic Engineering, with Applications to Network Revenue Management, *Proceedings of IEEE INFOCOM2003*, 2003
13. V. Rotar, *Probability Theory*, World Scientific, River Edge, NJ, 1997.
14. W. Ogryczak and A. Ruszczynski, Dual stochastic dominance and related mean–risk models, *SIAM Journal on Optimization*, vol. 13 pp. 60–78, 2002.
15. N. Semret, R. R.-F. Liao, A. T. Campell, and A. A. Lazar, Peering and Provisioning of Differentiated Internet Services, *Proceedings of IEEE INFOCOM2000*, pp. 100-107, 2000.
16. Y. Vardi, Network Tomography: Estimating Source-Destination Traffic Intensities from Link Data, *Journal of the American Statistical Association, Theory and Method*, **91**, pp. 365-377.