Stochastic Traffic Engineering for Demand Uncertainty and Risk-Aware Network Revenue Management

Debasis Mitra Qiong Wang
Bell Laboratories, Lucent Technologies, Murray Hill NJ 07974
{mitra,chiwang}@lucent.com

Abstract— **We present a stochastic traffic engineering framework for optimizing bandwidth provisioning and route selection in networks. The objective is to maximize revenue from serving demands, which are uncertain and specified by probability distributions. We consider heterogenous demands with different unit revenues and uncertainties. Based on mean-risk analysis, the optimization model enables a service provider to maximize mean revenue and contain the risk that the revenue falls below an acceptable level. Our framework is intended for off-line traffic engineering design, which takes a centralized view of network topology, link capacity, and demand. We obtain conditions under which the optimization problem is an instance of convex programming and therefore efficiently solvable. We also study the properties of the solution and show that it asymptotically meets the stochastic efficiency criterion. We derive properties of the optimal solution for the special case of Gaussian distributions of demands. We focus on the impact of demand uncertainty on various aspects of traffic engineering, such as link utilization, bandwidth provisioning and total revenue. The service provider's tolerance to risk is shown to have a strong influence on traffic engineering and revenue management decisions. We develop the** ***efficient frontier*****, which is the entire set of Pareto optimal pairs of mean revenue and revenue risk, to aid the service provider in selecting an appropriate operating point.**

Keywords- **Mathematical Programming, Economics, Traffic Engineering, Demand Uncertainty, Risk**

I. Introduction

Traffic engineering (TE) is a mechanism for traffic and revenue management in networks [5]. Supported by emerging technologies, especially Multi-Protocol Label Switching (MPLS), TE performs provisioning and admission control functions to optimize network operators' objectives [4], [10], [8]. The TE mechanism takes two complementary forms, on-line and off-line [5], [32]. On-line TE responds to changes of the network state in real-time. See [15] and [30] for works where the focus is on distributed on-line traffic engineering and provisioning. Off-line TE applies on a longer time-scale. Instead of focusing on instantaneous network states and individual connections, it takes as input statistics of aggregated traffic demands between node pairs. Combining this demand information with a centralized view of network topology and link capacites, off-line TE selects the topology of routes and provisions resources on the selected routes for carrying the demands. These decisions are optimized globally for demands of various service types and origin-destination pairs [3], [24], [26], [31]. The solution of the off-line optimization has been proposed as a reference point for on-line operations. For example, in [31] capacity preallocated by the off-line TE process is used as the threshold for on-line admission control. Similarly, in [9] the off-line provisioning process is used to guide the real-time routing and admission control.

The authors acknowledge K. G. Ramakrishnan for his contributions to the early development of this work, and Alan Weiss for the benefit of discussions.

This paper focuses on the optimization of off-line traffic engineering. 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. In previous work, the problem has been formulated as a deterministic multicommodity flow (MCF) model, where demand for each service and node pair is given as a fixed quantity, such as the expected value of forecasted traffic load [24], [31]. The goal is to find an appropriate amount of traffic to admit for each demand, and capacitated route(s) to carry the traffic, so that the service provider's objective, usually formulated as revenue earned by serving demands, is optimized subject to network capacity constraints.

This paper develops a stochastic traffic engineering framework for decision making that takes into account demand uncertainty and its consequence, namely, risk. The overall objective is to employ traffic engineering and revenue management techniques to achieve robustness in network performance and resulting revenue. This framework uses probabilistic distributions of demands as inputs for off-line optimization. Such distributional information is typically a byproduct of statistical procedures for forecasting network traffic from measurements [33], [6], [7], [23]. However, such information has not been used extensively in the past for lack of a suitable modeling framework.

The framework of this paper remedies several shortfalls of the deterministic approach to traffic engineering. For instance, in the deterministic MCF model, revenue derived from carrying demand is assumed to increase linearly with the amount of provisioned capacity up to the point where all traffic demand is satisfied. When demand is random, if more capacity is provisioned then the probability that the incremental capacity is fully utilized decreases. Consequently, mean revenue increases with a decreasing rate with the amount of provisioned capacity. This non-linear effect is captured in the stochastic traffic engineering framework. Incidentally, this also illustrates one of the reasons why the latter framework is intrinsically non-linear. Importantly, the framework also allows the variability of uncertainty in the traffic demand to be taken into account, and its impact assessed. Even when mean demands stay fixed, increasing uncertainty significantly affects the optimal traffic engineering solution. Typically the consequences include lower link utilizations and higher link shadow costs. Note that optimal selections of path topologies and provisioning of resources are based on link shadow costs [26], hence uncertainty impacts route selection and provisioning.

Uncertainty breeds risk. A service provider must concern itself not only with mean revenue and the strategies for its maximization, but also with the risk of revenue falling below acceptable levels. This risk is not calculable in the deterministic traffic engineering framework. Calculating risk and managing it are important components of network revenue management, and also of this work. The framework here allows the service provider to trade-off different objectives in its decision-making, such as the maximization of mean revenue against minimization of the risk of revenue shortfall.

The objective function in the optimization model of 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 [13], [21]. Robust engineering extends the mean-risk analysis to areas like power capacity expansion, aircraft scheduling and structural design [25]. Classical models of revenue and resource management, such as the Newsvendor problem, incorporate demand uncertainty. Recent work has extended these models to address risk aversion in the expected utility framework [1]. Risk in networking has its own characteristics. At the most basic level, demands of various service types between node pairs are associated with varying degrees of uncertainty; these demands compete for fixed network resources. The allocation of network resources to demands is also subject to servce providers' attitude towards risk, some service providers being more risk-averse than others. Taking the service provider's risk aversion into account in decision-making for network revenue management is clearly important, and this is done in the framework of this paper.

Considerable attention is given in this paper to finding a measure of risk that is appropriate specifically for network revenue management in the mean-risk approach. There are several candidates for such a risk index. Some of these candidates, such as the variance of revenue, lead to inferior solutions that are stochastically dominated by other feasible solutions in the context of our problem. 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 make the case that the standard deviation of network revenue as a measure of risk is an attractive compromise between stochastic efficiency and tractability.

The framework in the paper is sufficently general to include demands with different degrees of uncertainty, even for the same (source,destination) pair. Service providers routinely complement retail services by wholesale contracts with large customers for guaranteed demands at discounted prices; for example, see [17] and [18] regarding Level 3's wholesale agreement with Microsoft. Our framework is used to model the service provider's decision-making in allocating bandwidth to demands that range from highly-priced, but uncertain, to lower-priced, but with little or no uncertainty. A proper mix of these types of demands allows the service provider to mitigate revenue risk while maximizing expected revenue. This capability of our model is demonstrated numerically in Section VI. We recognize that the volume of guaranteed demands depends on the price discount. However, the scope of this paper does not allow us to address price-demand relationships. Hence, in our numerical examples, we assume that the price of guaranteed demand is heavily discounted, and correspondingly the volume is abundant.

As another aid to decision-making, we develop the efficient frontier, which is the complete set of Pareto optimal pairs of mean revenue and revenue risk, and thus is the totality of all rational solutions that the service provider needs to consider. As we will see in Section VI, the shape of the efficient frontier provides valuable information to the service provider in selecting the appropriate operating point.

With natural extensions, the framework of this paper has the potential of contributing to studies of 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 resource sharing among all or a subset of these service providers will benefit not only the participating service providers but society and consumers also. Modeling frameworks will be needed to quantify the value proposition from such bandwidth sharing agreements. It is possible to envisage service providers being engaged in dynamic, game-theoretic iterations of actions and reactions, in which basic moves of each player are computed by an extension of the present model. Also, our framework is potentially useful in service provider's decision-making of optimal investments in infrastructure over multi-year planning horizons in the presence of technological innovations and demand uncertainty. We believe that the framework presented here can be enhanced to handle such challenges.

The paper is organized as follows: In Section II we formulate the stochastic traffic engineering problem and present the optimization model. In Section III we discuss modeling of the risk term in the objective function and in Section IV we analyze the complexity of obtaining the global optimal solution. In Section V we derive properties of the optimal solution that provide important insights on bandwidth provisioning and route selection. In Section VI we give numerical results that show the impact of demand uncertainty and service provider's risk tolerance on network traffic engineering and revenue management. We also develop the efficient frontier. We present our concluding remarks in Section VII.

II. Model

A. Conception

A.1 Demand, Provisioning, and Revenue Management

In this paper, we define demand as the aggregated amount of traffic volume that customers are willing to pay the service provider to deliver from one network node to another. The price per unit of delivered traffic is assumed to be given. It is sometimes useful to define multiple demands for the same node pair, with the demands distinguished by the unit price, quality of service requirements, and volume statistics. For example, in [27] and [28], for the same end nodes, there is a distinction between retail demand, which generates higher unit revenue but for which the volume is subject to uncertainty, and wholesale demand for which the volume is known with certainty.

In the case of deterministic traffic engineering, the value of a demand is characterized by its unit price. In the presence of demand uncertainty, unit price alone is not sufficient to characterize the value of a demand. It is possible that a demand of high unit price is very volatile, in which case it is necessary to take into account the probability that demand volume will fall below the provisioned bandwidth. Consequently, the service provider should anticipate the possibility of a smaller revenue from such demands than from demands with lower unit prices but guaranteed volume. To optimally provision bandwidth in our framework, unit price should be used in conjunction with distributional information of demands to characterize the mean revenue return and revenue risk from bandwidth provisioning.

Revenue management under demand uncertainty is further complicated by the service provider's attitude towards risk. For example, a provisioning decision that guarantees a revenue of $99 may be more favorable to some service providers than the one that generates revenue of $10000 with probability 0.01 and zero revenue with probability 0.99, though the latter yields a higher mean revenue. In less extreme cases, the service provider may still want to give up a certain amount of mean revenue in exchange for reduced level of revenue risk. The service provider's attitude towards risk is influenced by several factors, such as credit ratings, cash reserves, and strategic needs to break into new markets. The service provider's risk tolerance should be reflected in a risk-aware revenue management model, as is the case in this paper.

We digress to observe that in the model here prices are not influenced by capacity provisioning decisions. It is possible to couple the two through price-demand relationships, as in [26], [19], for instance. The value in these models is in decision-making over time scales that are longer than is of interest in this paper.

A.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. 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 a switching node and consequent incremental delay. The admissibility of a route may also depend on policy, which typically reflects diverse considerations, such as security, the capability of switching nodes in the routes to handle certain services, etc. Generating the admissible route sets is a substantial task in itself. It is assumed in this paper that these sets are given.

B. Model Formulation

We formulate the network as a collection of nodes and links $(\mathcal{N}, \mathcal{L})$, where link $l \in \mathcal{L}$ has bandwidth C_l. Let $\mathcal{V}$ be the set of all demands and denote demand volume by $T_v (v \in \mathcal{V})$. In a restricted case where there is only one demand between each pair of network nodes, $\mathcal{V}$ corresponds to a subset of all node pairs. However, in this paper, each source-destination pair is allowed to have multiple demands, in which case $v \in \mathcal{V}$ is indexed by both node pair and demand type. Different demand types for the same node pair may be characterized by the unit price, admissible route set, GoS requirement, as well as traffic volume distribution that reflects the degree of uncertainty.

Let demand volume T_v be a random variable characterized by its Probability Density Function (PDF) and Cumulative Distribution Function (CDF), denoted by $f_v(x)$ and $F_v(x)$, respectively. Let $d_v (v \in \mathcal{V})$ be the amount of capacity provisioned to demand v. The provisioned quantity, d_v, may be routed on one or more admissible routes. Denote the admissible route set for $v \in \mathcal{V}$ by $\mathcal{R}(v)$ and let $\xi_r (r \in \mathcal{R}(v))$ be the amount of capacity provisioned on route r. Then

$$d_v = \sum_{r \in \mathcal{R}(v)} \xi_r. \tag{1}$$

It follows that the amount of carried demand

$$x_v(d_v) = \min(T_v, d_v). \tag{2}$$

Let $m_v(d_v)$ and $s_v^2(d_v)$ be the mean and variance of x_v, respectively. Then

$$m_v(d_v) = \int_0^{d_v} x f_v(x)dx + d_v \bar{F}_v(d_v) = \int_0^{d_v} \bar{F}_v(x)dx \tag{3}$$

and

$$\begin{aligned} s_v^2(d_v) &= \int_0^{d_v} x^2 f_v(x)dx + d_v^2 \bar{F}_v(d_v) - m_v^2(d_v) \\ &= 2\int_0^{d_v} x\bar{F}_v(x)dx - m_v^2(d_v), \end{aligned} \tag{4}$$

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

$$\begin{aligned} \frac{\partial m_v}{\partial d_v} &= \bar{F}_v(d_v) \geq 0, \\ \frac{\partial s_v^2}{\partial d_v} &= 2[d_v - m_v(d_v)]\bar{F}_v(d_v) \geq 0, \end{aligned} \tag{5}$$

i.e., both the mean and variance of carried demand increase with the amount of bandwidth provisioned. Their maximum values, denoted by $m_v(\infty)$ and $s_v^2(\infty)$, are the mean and variance of the demand v, respectively.

It is worth noting that the model for deterministic traffic engineering is a special case of the model above. In the former case, network demand is given by the 'demand matrix' of constants $\bar{T}_v$. We formulate the demand matrix as an instance of our model by using the following degenerate distribution function,

$$F_v(x) = \begin{cases} 0, & x < \bar{T}_v \\ 1, & x \geq \bar{T}_v \end{cases}.$$

In this case, from (3) and (4),

$$m_v(d_v) = \int_0^{d_v} \bar{F}_v(x)dx = \min(\bar{T}_v, d_v) \tag{6}$$

and

$$s_v^2(d_v) = 2\int_0^{d_v} x\bar{F}_v(x)dx - m_v^2(d_v) = 0. \tag{7}$$

Our model also accommodates mixed cases where both deterministic and random demands exist [27],[28].

Given fixed unit revenue, π_v, the revenue derived from serving demand v equals $\pi_v x_v(d_v)$, for which the mean is $\pi_v m_v(d_v)$ and the variance is $\pi_v^2 s_v^2(d_v)$. The total revenue is

$$W = \sum_{v \in \mathcal{V}} \pi_v x_v(d_v). \tag{8}$$

In the general case, $x_v(d_v)$ is a random variable for at least some $v \in \mathcal{V}$, and so is W. The service provider's objective function, denoted by Θ, is itself a function of W. The composition of Θ should reflect the service provider's desire to increase mean revenue and reduce the risk of revenue shortfall due to the randomness of W. In the next section, we discuss in detail the specification of Θ that satisfies this objective.

The overall optimization model is as follows:

$$\max \Theta(W(d_v, \xi_r)) \tag{9}$$

subject to:

$$\sum_{r \in \mathcal{R}(v)} \xi_r = d_v \ (v \in \mathcal{V}) \tag{10}$$

$$\sum_{r \in \mathcal{R}(v): l \in r} \xi_r \leq C_l \qquad l \in \mathcal{L} \tag{11}$$

$$0 < \underline{d}_v \leq d_v \ (v \in \mathcal{V}), \ \ 0 \leq \xi_r \ (r \in \mathcal{R}(v)), \tag{12}$$

The constraints in (10) specify bandwidth provisioning for demands on admissible routes, and (11) limit the total amount of provisioned bandwidth from exceeding link capacities. The lower bounds (12) on the provisioned bandwidth deserve additional comment. The parameter $\underline{d}_v$ is defined as the minimum bandwidth that must be provisioned for demand v. The parameter value is determined by factors such as regulation, service level agreements, and customer expectations on the grade of service.

Notice that with fixed link capacities, it may be infeasible to satisfy the minimum bandwidth requirement of some demands. Should this situation arise, the service provider has to supplement link capacities by reconfigurations or by buying bandwidth from other service providers. Our model can be extended to cover this scenario in the following way. Let p_l be the price of buying unit bandwidth on link l, and b_l be the amount to buy, which is a decision variable [28]. We can integrate buying decisions into the model by generalizing the revenue function in (8) to be

$$\hat{W} = W - \sum_{l \in \mathcal{L}} p_l b_l,$$

and replacing W with $\hat{W}$ in the objective function (9). Constraints in (11) should also be changed to

$$\sum_{r \in \mathcal{R}(v): l \in r} \xi_r \leq C_l + b_l \qquad l \in \mathcal{L}. \tag{13}$$

The generalized model makes it always feasible to provide minimum quantity guarantee for demands and its output recommends the optimal amount of bandwidth to buy on each link. Nevertheless, the generalized model that incorporates buying decisions is not the focus of this paper. Consequently, we will analyze the original model in (9)-(12), and assume that a feasible solution exists for the existing link capacities.

III. Modeling Risk

In this section, we take uncertainty in demand into consideration in the composition of an objective function that reflects both the maximization of mean revenue and the containment of the risk of revenue shortfall. We review two relevant risk modeling frameworks in Section III-A and discuss our characterization of revenue risk in Section III-B.

A. Relevant Frameworks for Risk Modeling

A.1 Mean-Risk Model

Mean-risk analysis addresses the issue of risk averseness by offering a broader optimization objective. The approach starts by developing a risk index, which is a quantitative measure of the risk of revenue shortfall, based on the revenue distribution. It then maximizes the weighted combination of the mean revenue 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 revenue to avoid risk.

A.2 Stochastic Dominance

Stochastic dominance theory defines a partial ordering of random variables based on their probability distributions [16]. Let W_1 and W_2 be two random variables, which represent revenues 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* revenue target w, i.e.,

$$\forall w, \ Pr(W_1 \geq w) \geq Pr(W_2 \geq w). \tag{14}$$

Furthermore, W_1 stochastically dominates W_2 to the second degree iff

$$\forall w, \ \int_w^\infty Pr(W_1 \geq \zeta) d\zeta \geq \int_w^\infty Pr(W_2 \geq \zeta) d\zeta. \tag{15}$$

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

B. Formulation of the Objective Function

We formulate the objective function as an instance of the mean-risk model and require that the solution that optimizes it be stochastically efficient. Whether these conditions can be met depends on the characteristics of the revenue distribution and the choice of the risk index. In portfolio management, variance is commonly used as the risk index and the objective is to maximize (mean - $\delta *$variance) [16], However, as we show in Section III-B.1, in the context of network revenue management, the mean-variance model often leads solutions that are stochastically dominated. On the other hand, applying other risk measures that guarantee stochastic efficiency, such as Tail Value at Risk defined in Section III-B.2, lead to models that are too complex to solve. As a compromise, in Section III-B.3 we propose standard deviation as the risk index.

B.1 Variance

Variance of network revenue is a natural candidate for the risk index, especially since it has been widely used in many mean-risk models. The objective function that uses the variance as the risk index is

$$\Theta = E(W) - \delta Var(W) = \sum_{v \in \mathcal{V}} \pi_v [m_v(d_v) - \delta \pi_v s_v^2(d_v)] \tag{16}$$

Notice that (16) reflects the aforementioned assumption that demands between different node pairs are independent. Applying (3) and (4),

$$\frac{\partial \Theta}{\partial d_v} = \bar{F}_v(d_v)[1 - 2\delta \pi_v (d_v - m_v)]$$

$$= \bar{F}_v(d_v)[1 - 2\delta\pi_v \int_0^{d_v} F_v(x)dx]. \quad (17)$$

The equation $2\delta\pi_v \int_0^{d_v} F_v(x) = 1$ has an unique solution, $\bar{d}_v$, and this solution has the following properties:

$$\frac{\partial\Theta}{\partial d_v} \geq 0,\ \frac{\partial^2\Theta}{\partial d_v^2} \leq 0 \text{ if } d_v \leq \bar{d}_v \text{ and } \frac{\partial\Theta}{\partial d_v} < 0,\ \text{if } d_v > \bar{d}_v. \quad (18)$$

Based on this observation, we impose $\max(\underline{d}_v, \bar{d}_v)$ as an upper bound on d_v, which makes our model a concave maximization problem that can be solved efficiently. If $\bar{d}_v \leq \underline{d}_v$, then $d_v = \underline{d}_v$ by (12). Otherwise, $d_v \leq \bar{d}_v$, which is a new constraint that reduces the original feasible region (a polyhedron defined by (10), (11), and (12)) to a convex set on which the objective function (16) is concave. This additional restriction on d_v does not affect the optimal solution since by (18), increasing d_v beyond $\bar{d}_v$ only reduces the value of Θ.

We now show that the optimal solution of the mean-variance model can be stochastically dominated. Since the optimal value of d_v is bounded by $\max(\underline{d}_v, \bar{d}_v)$, when the network has more capacity than this bound for every demand, some links will have unused bandwidth in the optimal solution. In this case, a new feasible solution can be devised by increasing the bandwidth provisioned to node pairs that are connected by links with unused bandwidth, while bandwidths provisioned to all other node pairs are left unchanged. The latter solution will certainly generate no less revenue in every realization of demand, and will generate more revenue in cases where actual demand is greater than the aforementioned bound, $\max(\underline{d}_v, \bar{d}_v)$. Therefore, it dominates the optimal solution to the mean-variance model to the first degree. Observe that the phenomenon is likely to occur in real-world networks.

B.2 Tail Value at Risk

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

$$TVaR(p) = \int_0^p q_W(\eta)d\eta, \quad (19)$$

where $q_W(p) = \inf\{w | Pr(W \leq w) \geq p\}$ is the p-quantile of revenue W [16], [29].

We do not use TVaR as the risk index in our model because its selection makes optimization computationally difficult.

B.3 Standard Deviation as Risk Index

Our choice of risk index is the standard deviation of network revenue, $s(W)$. It 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 difference between the optimal solutions where the risk index is TVaR and the standard deviation, respectively, goes to zero as the number of node pairs 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 (8), W is a summation of many *independent* random revenues, $\pi_v x_v(d_v)$. From (5) the variance of each revenue, $\pi_v^2 s_v^2(d_v)$, monotonically increases with d_v, and thus has a lower bound $\pi_v^2 s_v^2(\underline{d}_v)$ from (12), and an upper bound $\pi_v^2 s_v^2(+\infty) = \pi_v^2 var(T_v)$. Therefore, for every feasible solution

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

i.e., Lindeberg's condition is satisfied, and the Central Limit Theorem can be applied as follows,

$$\begin{aligned}\frac{W - E(W)}{s(W)} &= \frac{\sum_{v \in \mathcal{V}} \pi_v [x_v(d_v) - m_v(d_v)]}{\sqrt{\sum_{v \in \mathcal{V}} \pi_v^2 s_v^2}} \\ &\to N(0,1), \quad \text{as } |\mathcal{V}| \to +\infty. \quad (20)\end{aligned}$$

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

$$\begin{aligned}TVaR(p) &= \int_0^p q_W(\eta)d\eta = \int_{-\infty}^{q_W(p)} x dF_W(x) \\ &\sim \int_{-\infty}^{q_N(p)} [E(W) + s(W)\theta] dF_N(\theta) \\ &= p[E(W) - \frac{e^{-q_N^2(p)}}{p\sqrt{2\pi}/2} s(W)], \quad (21)\end{aligned}$$

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 a unique inverse $p(\delta)$. By (21), the solution that maximizes $E(W) - \delta s(W)$ also maximizes $TVaR(p)$ where $p = p(\delta)$, and thus is stochastically efficient.

IV. Model Analysis and Properties

The preceding section establishes that the standard deviation is an appropriate choice of the risk index in large networks. In this section we demonstrate that the use of standard deviation also leads to a tractable optimization model. We also discuss the implicattions of the necessary conditions for optimality.

Using the standard deviation as the risk index, the objective function in (9) becomes

$$\max \Theta(W(d_v,\xi_r)) = \sum_{v\in\mathcal{V}} \pi_v m_v(d_v) - \delta\sqrt{\sum_{v\in\mathcal{V}} \pi_v^2 s_v^2(d_v)}. \tag{22}$$

Given that all constraints of the above model are linear, the difficulty of finding the global optimum depends on the shape of the objective function Θ. If Θ is concave in all nonlinear variables $d_v (v \in \mathcal{V})$, the model falls into the class of concave maximization problems. In this case, the global optimum can be found efficiently with existing standard algorithms.

In general, Θ is not concave everywhere if $\delta > 0$, as can be verified by considering a restricted case with only one nonlinear variable. Despite this inconvenience, we show in this section that in many circumstances, the model can still be solved as a concave maximization problem. Our approach is based on two theorems developed in IV-A and the subsequent analysis in IV-B.

A. The Shape of the Objective Function

In the following, we use m_v and s_v^2 to represent $m_v(d_v)$ and $s_v^2(d_v)$, as defined in equations (3) and (4), respectively. We also omit the argument d_v in the distribution functions, and use F_v, $\bar{F}_v$, and f_v to represent $F_v(d_v)$, $\bar{F}_v(d_v)$, and $f_v(d_v)$, respectively. Lemma 1 is the basis for Theorems 1 and 2.

Lemma 1 For any $v \in \mathcal{V}$ and $d_v \geq 0$,

$$F_v s_v^2 \geq (d_v - m_v)^2 \bar{F}_v. \tag{23}$$

Proof: Because $f_v \geq 0$, from equations (3) and (4):

$$\begin{aligned}\frac{\partial (F_v s_v^2)}{\partial d_v} &= 2F_v\bar{F}_v(d_v - m_v) + f_v s_v^2 \\ &\geq 2F_v\bar{F}_v(d_v - m_v) - f_v(d_v - m_v)^2 \\ &= \frac{\partial[(d_v - m_v)^2\bar{F}_v]}{\partial d_v}.\end{aligned} \tag{24}$$

Since $F_v s_v^2 = (d_v - m_v)^2\bar{F}_v = 0$ at $d_v = 0$,

$$F_v s_v^2 \geq (d_v - m_v)^2\bar{F}_v \quad \text{for all} \quad d_v \geq 0. \quad \square \tag{25}$$

We show in Theorem 1 that Θ is unimodal in every d_v.

Theorem 1 For $v \in \mathcal{V}$, given fixed values for $d_{v'} (v' \neq v)$,

$$\frac{\partial\Theta}{\partial d_v} = \begin{cases} \geq 0 & \text{if } 0 \leq d_v \leq \hat{d}_v \\ < 0 & \text{if } d_v > \hat{d}_v \end{cases}, \tag{26}$$

$$\text{where} \quad \hat{d}_v = \sup\{d_v : \frac{(d_v - m_v)^2}{s_v^2 + \Psi_v} \leq \frac{1}{\delta^2}\}. \tag{27}$$

Here $\Psi_v = \sum_{v'\neq v}(\pi_{v'}/\pi_v)^2 s_{v'}^2$, and $\hat{d}_v$ is the unique solution to (27), i.e., $\hat{d}_v$ is calculated by solving the equation obtained by replacing $\leq$ by $=$.

Proof: Let $S^2 = Var(W) = \sum_{v\in\mathcal{V}} \pi_v^2 s_v^2$,

$$\begin{aligned}\frac{\partial\Theta}{\partial d_v} &= \pi_v\frac{\partial m_v}{\partial d_v} - \delta\frac{\partial S}{\partial d_v} = \pi_v\bar{F}_v - \delta\pi_v\bar{F}_v\frac{(d_v - m_v)}{S/\pi_v} \\ &= \pi_v\bar{F}_v[1 - \delta\frac{(d_v - m_v)}{\sqrt{s_v^2 + \Psi_v}}].\end{aligned} \tag{28}$$

Notice that by Lemma 1,

$$\partial[\frac{(d_v - m_v)}{\sqrt{s_v^2 + \Psi_v}}]/\partial d_v = \frac{F_v(s_v^2 + \Psi_v) - (d_v - m_v)^2\bar{F}_v}{\sqrt{(s_v^2 + \Psi_v)^3}} \geq 0. \tag{29}$$

Therefore $(d_v - m_v)/\sqrt{s_v^2 + \Psi_v}]$ monotonically increases from 0 to $+\infty$ as d_v goes from 0 to $+\infty$. It follows that $\hat{d}_v$ as defined by (27) is unique, and $\partial\Theta/\partial d_v >, =$, or < 0, depending on whether $d_v <, =$, or $> \hat{d}_v$. $\square$

Holding (27) at equality and applying the Implicit Function Theorem,

$$\frac{\partial\hat{d}_v}{\partial\Psi_v} = \frac{d_v - m_v}{F_v(s_v^2 + \Psi_v) - (d_v - m_v)^2\bar{F}_v} \geq 0, \tag{30}$$

indicating that $\hat{d}_v$ increases with Ψ_v. This result will be used in IV-B.

Clearly, any maximum point of Θ can be reached only in areas where $d_v \leq \hat{d}_v$ for all $v \in \mathcal{V}$. Theorem 2 shows that Θ is concave in this region. Before presenting the theorem, we first give the second-order derivatives,

$$\begin{aligned}\frac{\partial^2\Theta}{\partial d_v^2} &= -\pi_v f_v[1 - \delta\frac{(d_v - m_v)}{S/\pi_v}] \\ &\quad -\delta\frac{\pi_v\bar{F}_v}{S/\pi_v}[F_v - (\frac{(d_v - m_v)}{S/\pi_v})^2\bar{F}_v] \quad v \in \mathcal{V},\end{aligned} \tag{31}$$

and for $v \neq v'$,

$$\frac{\partial^2\Theta}{\partial d_v \partial d_{v'}} = \frac{\delta}{S^3}\pi_v^2\pi_{v'}^2(d_v - m_v)\bar{F}_v(d_{v'} - m_{v'})\bar{F}_{v'} \tag{32}$$

Theorem 2 : Let $H(\Theta)$ be the Hessian matrix of Θ. If

$$\frac{\partial\Theta}{\partial d_v} = \pi_v\bar{F}_v[1 - \delta\frac{(d_v - m_v)}{S/\pi_v}] \geq 0 \text{ for all } v \in \mathcal{V} \tag{33}$$

then $H(\Theta)$ is negative semi-definite.

Proof:

$$H(\Theta) = \begin{pmatrix} H_1 & 0 \\ 0 & 0 \end{pmatrix} \tag{34}$$

where H_1 is a square matrix of dimension $n = |\mathcal{V}|$, and its entries are $\partial^2\Theta/\partial d_{v_i}\partial d_{v_j}(v_i, v_j \in \mathcal{V})$. Other elements that take the zero value correspond to second-order derivatives with respect to linear variables in the objective function. $H(\Theta)$ is negative semi-definite if and only if H_1 is negative semi-definite.

Apply equations (31) and (32), and note that $S^2 = \pi_v^2 s_v^2 + \sum_{v'\neq v} \pi_{v'}^2 s_{v'}^2$ for any $v \in \mathcal{V}$, we have

$$H_1 = -H_{1a} - \frac{\delta}{S^3}H_{1b} - \frac{\delta}{S^3}H_{1c}, \tag{35}$$

where elements of matrix H_{1a}, H_{1b}, and H_{1c}, denoted correspondingly as $h_a(i,j)$, $h_b(i,j)$, and $h_c(i,j)$. H_{1a} and H_{1b} are diagonal matrices with

$$h_a(i,i) = \pi_{v_i} f_{v_i}[1 - \delta\frac{(d_{v_i} - m_{v_i})}{S/\pi_{v_i}}], \text{ and}$$
$$h_b(i,i) = \pi_{v_i}^4 \bar{F}_{v_i}[s_{v_i}^2 F_{v_i} - (d_{v_i} - m_{v_i})^2 \bar{F}_{v_i}]. \tag{36}$$

Also, $h_c(i,i) = \pi_{v_i}^2 \bar{F}_{v_i} F_{v_i}(\sum_{v\neq v_i} \pi_v^2 s_v^2)$, and for $i \neq j$,

$$h_c(i,j) = -\pi_{v_i}^2 \pi_{v_j}^2 (d_{v_i} - m_{v_i})\bar{F}_{v_i}(d_{v_j} - m_{v_j})\bar{F}_{v_j}. \tag{37}$$

H_{1a} is positive semi-definite by assumption, and H_{1b} is positive semi-definite by Lemma 1. To prove that $H(\Theta)$ is negative semi-definite, it suffices to show that H_{1c} is positive semi-definite, i.e., for any real vector $\vec{X}$ of dimension $n = |\mathcal{V}|$,

$$\vec{X}^T H_{1c}\vec{X} = \sum_{v\in\mathcal{V}} x_v^2 \pi_v^2 \bar{F}_v F_v(\sum_{v'\neq v} \pi_{v'}^2 s_{v'}^2)$$
$$- \sum_{v'\neq v} x_v x_{v'} \pi_v^2 \pi_{v'}^2 (d_v - m_v)\bar{F}_v(d_{v'} - m_{v'})\bar{F}_{v'} \geq 0. \tag{38}$$

This is true because by Lemma 1,

$$\sqrt{\bar{F}_v F_v \bar{F}_{v'} F_{v'} s_v^2 s_{v'}^2} \geq (d_v - m_v)\bar{F}_v(d_{v'} - m_v)\bar{F}_{v'}, \tag{39}$$

and by the A-G Mean Inequality(i.e., $a^2 + b^2 \geq 2ab$),

$$\sum_v x_v^2 \pi_v^2 \bar{F}_v F_v(\sum_{v'\neq v} \pi_{v'}^2 s_{v'}^2)$$
$$\geq \sum_{v'\neq v} \pi_v^2 \pi_{v'}^2 x_v x_{v'} \sqrt{\bar{F}_v F_v \bar{F}_{v'} F_{v'} s_v^2 s_{v'}^2}. \quad \square \tag{40}$$

B. Discussions of Properties

The two theorems in the last section define a set:

$$\Omega \equiv \{d_v(v \in \mathcal{V}) : \underline{d}_v \leq d_v \leq \hat{d}_v\}, \tag{41}$$

over which Θ is concave (if $\underline{d}_v \geq \hat{d}_v$, then d_v is fixed at $\underline{d}_v$). The set Ω also contains all local maximum point(s) of Θ. The model is a concave maximization problem if Ω, or its intersection with the feasible region, is a convex set. In this case, we can find the global optimum efficiently.

The set Ω may or may not be convex, depending on demand distribution, parameters $\underline{d}_v$ and δ. To verify that Ω may not be convex, consider a single-link network that serves two demands. Assume that volumes of both demands are uniformly distributed over $[0, 1]$, and the ratio of their unit prices $\pi_1/\pi_2 = 2$. Let $\delta = 1$ and $\underline{d}_1 = \underline{d}_2 = 0.1$. From (27) and (41), Ω is the set of (d_1, d_2) that satisfies

$$d_1 \geq 0.1, \; d_2 \geq 0.1, \text{ and}$$
$$3d_1^4 - 2d_1^3 \leq 8d_2^3 - 6d_2^4, \;\; 6d_2^4 - 4d_2^3 \leq d_1^3 - 3d_1^4/4.$$

The set is not convex. For instance, it contains $(0.67, 0.11)$ and $(0.79, 0.33)$ but not their average $(0.73, 0.22)$.

In case Ω is not convex, we can still solve the model as a concave maximization problem under certain conditions. Consider $\hat{d}_v(v \in \mathcal{V})$ in Theorem 1, which are obtained by solving:

$$\frac{[\hat{d}_v - m_v(\hat{d}_v)]^2}{s_v^2(\hat{d}_v) + \Psi_v} = \frac{1}{\delta^2}, \text{ where } \Psi_v = \sum_{v'\neq v}(\frac{\pi_{v'}}{\pi_v})^2 s_{v'}^2(d_{v'}). \tag{42}$$

Since $\underline{d}_v > 0$ is the minimum amount of provisioned bandwidth for v, and $s_{v'}^2(d_{v'})$ increases with $d_{v'}$,

$$\bar{\Psi}_v = \sum_{v'\neq v}(\frac{\pi_{v'}}{\pi_v})^2 s_{v'}^2(\underline{d}_{v'})$$

is the lower bound of Ψ_v. Suppose that for each v, we solve (42) by using $\Psi_v = \bar{\Psi}_v$ and denote the solution by $\hat{d}_v^e$, i.e.,

$$\frac{[\hat{d}_v^e - m_v(\hat{d}_v^e)]^2}{s_v^2(\hat{d}_v^e) + \bar{\Psi}_v} = \frac{1}{\delta^2}. \tag{43}$$

Then $\hat{d}_v^e \leq \hat{d}_v$ because as mentioned in IV-A, $\hat{d}_v$ that solves (42) increases with Ψ_v. Consequently, the polyhedron

$$\Omega^e \equiv \{d_v(v \in \mathcal{V}) : \underline{d}_v \leq d_v \leq \hat{d}_v^e\}$$

is a subset of Ω. If Ω^e is large enough to contain the global optimum, the model can be solved by exercising a concave maximization algorithm on the polyhedron.

We now discuss conditions for the global optimum to be "trapped" inside Ω^e. Define

$$k_v = \frac{\bar{\Psi}_v}{s_{v,\infty}^2},$$

where $1/k_v$ is the ratio of the maximum variance of the revenue from node pair v to the minimum variance of the revenue from all other node pairs. Because $s_{v,\infty}^2 \geq s^2(\hat{d}_v^e)$,

$$\frac{1}{\delta^2} = \frac{[\hat{d}_v^e - m_v(\hat{d}_v^e)]^2}{s_v^2(\hat{d}_v^e) + \bar{\Psi}_v} \leq \frac{[\hat{d}_v^e - m_v(\hat{d}_v^e)]^2}{(k_v+1)s_v^2(\hat{d}_v^e)}.$$

By Lemma 1,

$$\frac{1}{\delta^2} = \frac{[\hat{d}_v^e - m_v(\hat{d}_v^e)]^2}{s_v^2(\hat{d}_v^e) + \bar{\Psi}_v} \leq \frac{F_v(\hat{d}_v^e)}{(k_v+1)\bar{F}_v(\hat{d}_v^e)},$$

so

$$F_v(\hat{d}_v^e) \geq \frac{k_v+1}{k_v+1+\delta^2}.$$

If $k_v \gg \delta^2$ for each v, then $F_v(\hat{d}_v^e) \approx 1$, and it becomes certain that Ω^e contains the global optimum. For example, when $k_v = 20$, $\delta = 1$, then $F_v(\hat{d}_v^e) \geq 0.95$. This means that if it is uneconomical to provision bandwidth for demand in excess of the 95% percentile, then the optimal solution can be obtained by concave maximization over Ω^e.

For the condition $k_v \gg \delta^2$ to hold, either δ is small (in an extreme case when $\delta = 0$, $\hat{d}_v^e = \infty$), or k_v is large. The latter corresponds to situations where the contribution of each node pair to the variance of total revenue is insignificant. This happens when the network has many node pairs, and the total revenue is not dominated by the revenue from an individul pair. This is typically the real-world case for large networks. For smaller networks, straightforward enumeration of extreme and boundary points is a viable option.

C. Necessary Conditions for Optimality and Implications

The Lagrangian of our model takes the following form:

$$\begin{aligned}\Lambda &= \sum_{v\in\mathcal{V}} \pi_v m_v - \delta\sqrt{\sum_{v\in\mathcal{V}} \pi_v^2 s_v^2} + \sum_{v\in\mathcal{V}} \chi_v\Big(\sum_{r\in\mathcal{R}(v)} \xi_r - d_v\Big) \\ &+ \sum_{l\in\mathcal{L}} \lambda_l\Big(C_l - \sum_{r\in\mathcal{R}(v):l\in r} \xi_r\Big) \qquad (44)\end{aligned}$$

It follows that the first-order necessary conditions are (using (28)):

$$\left.\begin{aligned}&\partial\Lambda/\partial d_v = \pi_v \bar{F}_v[1 - \delta\tfrac{\pi_v(d_v - m_v)}{S}] - \chi_v \leq 0 \\ &d_v(\partial\Lambda/\partial d_v) = 0,\ \ d_v \geq 0,\ \ \chi_v \geq 0\ \ (v \in \mathcal{V})\end{aligned}\right\} \quad (45)$$

$$\left.\begin{aligned}&\partial\Lambda/\partial\xi_r = \chi_v - \textstyle\sum_{l:l\in r} \lambda_l \leq 0 \\ &\xi_r(\partial\Lambda/\partial\xi_r) = 0,\ \xi_r \geq 0,\ \lambda_l \geq 0\ \forall r, l.\end{aligned}\right\} \quad (46)$$

The quantity λ_l is interpreted as the *link shadow cost*, which reflects the marginal value of capacity on link l and corresponds to dual variables in [2]. It is a critical quantity for route selection [2], for real-time routing and rate control [14], [35], and for off-line, joint design of pricing and provisioning [26].

Specifically,

1. By (46), for any route $r_0 \in \mathcal{R}(v)$, $\xi_{r_0} > 0$ only when

$$\sum_{l\in r_0} \lambda_l = \min_{r\in\mathcal{R}(v)} \sum_{l\in r} \lambda_l,$$

indicating that traffic is carried solely on the *shortest* path(s) of all the admissible routes, where link shadow cost, λ_l, is the distance metric.

2. We define $\chi_v = \min_{r\in\mathcal{R}(v)} \sum_{l\in r} \lambda_l$ as the *demand shadow cost*, i.e., the opportunity cost of carrying demand between $v \in \mathcal{V}$. By (45), the optimal quantity to be provisioned for demand v is determined at the point where the marginal increase of mean revenue, $\pi_v \bar{F}_v$, compensated by the marginal change of risk, $\delta\pi_v(d_v - m_v)/\sqrt{\sum_{v\in\mathcal{V}} \pi_v^2 s_v^2}$, equals the demand shadow cost, χ_v.

V. A Special Case: Truncated Gaussian Distribution

When demand between a node pair comes from many independent individual sources, the total demand can be approximated by the Gaussian distribution. The notion of Gaussian distribution for aggregated traffic demand between network nodes is well-established, specially in studies that extract traffic models from measurements and inference, such as tomographic, techniques. See for instance [22], [23], [6], [7]. We will make this assumption in what follows. Of course, the distribution needs to be restricted to nonnegative values, and the PDF should also be normalized properly, so that the total probability over the restricted sample space is unity. As a result, we will consider the *Truncated Gaussian Distribution* characterized by the following PDF:

$$f_v(x) = \frac{1}{\sqrt{2\pi}\sigma_v G_v} e^{-(x-\mu_v)^2/2\sigma_v^2} \ \ x \geq 0, \qquad (47)$$

where the normalizing parameter is:

$$G_v = \frac{Erfc(-\tau_v)}{2} \ \text{ and } \ \tau_v = \frac{\mu_v}{\sqrt{2}\sigma_v}, \qquad (48)$$

where $Erfc() = 1 - Erf()$.

In the rest of this paper, we will assume that μ_v is sufficiently larger than σ_v ($\mu_v > 2.75\sigma_v$), so that the truncation effect is negligible. In this case, μ_v and σ_v approximately equal the mean and standard deviation of the demand distribution, and are used as proxies for these parameters in the following analysis.

As before, d_v is the amount of bandwidth provisioned for demand v. Then the mean and standard deviation of carried demand may be derived to be

$$\begin{aligned} m_v(d_v) &= \frac{1}{\sqrt{2\pi}\sigma_v G_v}\int_0^\infty \min(x,d_v)e^{-(x-\mu_v)^2/2\sigma_v^2}dx \\ &= \mu_v + \sigma_v\gamma(\tilde{d}_v), \end{aligned} \tag{49}$$

$$\begin{aligned} s_v^2(d_v) &= \frac{\int_0^\infty \min(x^2,d_v^2)e^{-(x-\mu_v)^2/2\sigma_v^2}dx}{\sqrt{2\pi}\sigma_v G_v} - m_v^2 \\ &= \sigma_v^2[1-\frac{Erfc(\tilde{d}_v)}{2G_v} - \gamma^2(\tilde{d}_v) + \sqrt{2}\tilde{d}_v\gamma(\tilde{d}_v) \\ & -\frac{\tilde{d}_v+\tilde{\mu}_v}{\sqrt{\pi}G_v}e^{-\tilde{\mu}_v^2}], \end{aligned} \tag{50}$$

where $\tilde{d}_v = (d_v-\mu_v)/\sqrt{2}\sigma_v$, $\tilde{\mu}_v = \mu_v/\sqrt{2}\sigma_v$, and $\gamma(\tilde{d}_v) = [e^{-\tilde{\mu}_v^2} - e^{-\tilde{d}_v^2} + \sqrt{\pi}\tilde{d}_v Erfc(\tilde{d}_v)]/\sqrt{2\pi}G_v$.

The following inequalities provide important insights of the optimal solution.

Theorem 3 Let $F_v(x)$ be the CDF of the distribution specified by (47), Then

$$a) \quad \frac{\partial(m_v/d_v)}{\partial d_v} \le 0, \tag{51}$$

$$b) \quad \frac{\partial m_v}{\partial \sigma_v} \le 0 \text{ if } d_v \le 2\mu_v, \tag{52}$$

$$c) \quad \frac{\partial \bar{F}_v}{\partial \sigma_v} \ge 0 \text{ if } d_v \ge (1+\frac{e^{-\mu_v^2/2\sigma_v^2}}{G_v})\mu_v. \tag{53}$$

Proof: a) Since $\bar{F}_v(x)$ decreases in x, from (3),

$$\frac{\partial(m_v/d_v)}{\partial d_v} = \frac{d_v\bar{F}_v - \int_0^{d_v}\bar{F}_v(x)dx}{d_v^2} \le 0$$

b) From (3) and (47)

$$\begin{aligned} m_v &= \frac{1}{\sqrt{2\pi}G_v}[\int_{-\mu_v/\sigma_v}^{(d_v-\mu_v)/\sigma_v}(\sigma_v t+\mu_v)e^{-t^2/2}dt \\ &+ d_v\int_{(d_v-\mu_v)/\sigma_v}^{+\infty} e^{-t^2/2}dt] \end{aligned} \tag{54}$$

$$\frac{\partial m_v}{\partial \sigma_v} = \frac{e^{-\mu_v^2/2\sigma_v^2}}{\sqrt{2\pi}G_v}[1-e^{\frac{d_v(2\mu_v-d_v)}{2\sigma_v^2}}] - \frac{m_v}{G_v}\frac{\partial G_v}{\partial \sigma_v} \tag{55}$$

$$\frac{\partial G_v}{\partial \sigma_v} = \frac{\mu_v}{\sqrt{2\pi}\sigma_v^2}e^{-\mu_v^2/2\sigma_v^2} \ge 0, \text{ so} \tag{56}$$

$$\frac{\partial m_v}{\partial \sigma_v} \le 0 \text{ if } d_v \le 2\mu_v.$$

c) From (47)

$$\bar{F}_v = \frac{1}{\sqrt{2\pi}G_v}\int_{(d_v-\mu_v)/\sigma_v}^{\infty} e^{-t^2/2}dt \tag{57}$$

$$\begin{aligned} \frac{\partial \bar{F}_v}{\partial \sigma_v} &= \frac{(d_v-\mu_v)}{\sqrt{2\pi}G_v\sigma_v^2}e^{-(d_v-\mu_v)^2/2\sigma_v^2} \\ &- \frac{1}{\sqrt{2\pi}G_v^2}\frac{\partial G_v}{\partial \sigma_v}\int_{(d_v-\mu_v)/\sigma_v}^{+\infty} e^{-t^2/2}dt \end{aligned} \tag{58}$$

Because $e^{-x^2/2} > \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt$ for all $x \ge 0$,

$$\frac{\partial \bar{F}_v}{\partial \sigma_v} \ge \frac{e^{-(d_v-\mu_v)^2/2\sigma^2}}{G_v}[\frac{(d_v-\mu_v)}{\sqrt{2\pi}\sigma_v^2} - \frac{\partial G_v}{\partial \sigma_v}/G_v], \tag{59}$$

By (56), the right hand side equals

$$\frac{e^{-(d_v-\mu_v)^2/2\sigma^2}}{\sqrt{2\pi}\sigma_v^2 G_v}[d_v - (1+\frac{e^{-\mu_v^2/2\sigma_v^2}}{G_v})\mu_v], \tag{60}$$

which leads to the conclusion.□

Equation (51) shows that the ratio of the mean carried traffic to the provisioned bandwidth decreases with the increase of provisioned bandwidth. Notice that this trend of declining return from bandwidth provisioning is not specific to the truncated Gaussian distribution assumption, as can be seen from the proof.

Theorems 3-b) and 3-c) reveal an interesting implication of demand uncertainty, which is parameterized by the standard deviation σ_v. In (52), the condition $d_v \le 2\mu_v$ is imposed to accommodate the effect of truncating the demand distribution, and can be ignored if the distribution is close to Gaussian. The inequality shows that increasing the standard deviation of a demand reduces the mean carried traffic, indicating that demand uncertainty is detrimental to revenue. Does that mean that less bandwidth should be provisioned to a demand if its uncertainty increases? Not necessarily. It is implied by c) of Theorem 3 that when the demand is provisioned more bandwidth than its mean, which is typical, it is optimal to allocate more bandwidth to it as its uncertainty increases. To make this point, we refer to the necessary condition in (45), and consider the special case in which $\delta = 0$, so that

$$\pi_v\bar{F}_v = \chi_v.$$

Suppose that the condition in (53) is satisfied, so that $\partial\bar{F}_v/\partial\sigma_v \ge 0$. Then to maintain the above equality as σ_v increases, either d_v has to increase to cancel out the impact on $\pi_v\bar{F}_v$ by the increase of σ_v (notice that $\partial\bar{F}_v/\partial d_v \le 0$), or χ_v has to increase. The increase in χ_v leads to an increase in the shadow costs of some links on route $r \in \mathcal{R}(v)$. The latter effectively makes routes of other demands that share these links with v more expensive. As a result, the amount of bandwidth provisioned to these demands will be reduced, and hence more bandwidth will be left for demand v.

To summarize, the above discussion shows that while a demand of higher uncertainty usually gives lower mean revenue, it may also have a higher mean *marginal* revenue (represented by the term $\pi_v \bar{F}_v$), which is the incremental revenue obtained from an additional amount of provisioned bandwidth. This upward revenue potential may lead to more bandwidth being dedicated to high-uncertainty demands, despite the fact that the average revenue from these demands are low. This result is premised on the condition on d_v in (53) being satisfied, which requires the provisioned bandwidth to be no less than the mean demand, amplified slightly by the term $e^{-\mu_v^2/2\sigma_v^2}/G_v$ to accommodate the truncation of the Gaussian distribution. The requirement is usually satisfied in networks that are not over-loaded.

The service provider's attitude to risk influences the provisioning of bandwidth to demands with higher uncertainty. Referring to the necessary condition for optimality (45),

$$\pi_v \bar{F}_v [1 - \delta \frac{\pi_v (d_v - m_v)}{S}] = \chi_v,$$

in the discussion of the special case of $\delta = 0$ we showed that when σ_v increases, d_v also needs to be increased to rebalance the equation by reducing its left-hand side. In case $\delta \neq 0$, raising d_v not only reduces $\bar{F}_v$ but also increases $\pi_v(d_v - m_v)/S$. Therefore, the equation can be rebalanced with a smaller increase in d_v. This suggests that a service provider with a smaller risk tolerance will provide less bandwidth for demands with high uncertainty. This phenomenon is observed in our numerical examples, which are reported in the next section.

VI. Numerical Studies

In this section, we discuss implications of demand uncertainty and service provider's risk tolerance on optimal solutions for traffic engineering and revenue management through numerical examples. We first describe the network topology and base case scenario in subsection VI-A, and present results from the numerical investigations in subsection VI-B to VI-E.

A. Network and Base Case

We consider a sample network which has 12 nodes and 14 bidirectional links. The network topology, as well as the labels of nodes and links are shown in Figure 1. All links have 150 units of capacity, except links $6, 7, 12, 13$ and 14, which have 200 units. The latter links are given higher capacities since they are likely to carry more traffic on account of their central locations. Between each node pair in the network there is some random demand which

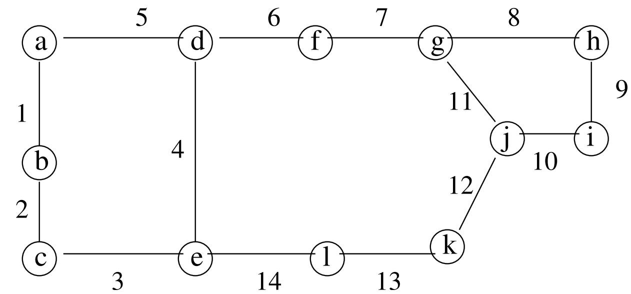

Fig. 1. Network Topology

is symmetric in both directions. We make the following assumptions about these demands.

1. Let h_v be the minimum number of hops between a node pair v. The admissible route set of demand v, R_v, is composed of all paths that have no more than $(h_v + 2)$ links.
2. The price per unit of carried bandwidth for demand v, π_v, is proportional to h_v, i.e., $\pi_v = \kappa h_v$, where κ is set at 50 in the numerical studies. Hence, the price for a given demand is independent of its selected routes.
3. The volume of each demand follows the Truncated Gaussian distribution, see (47). The ratio of the two defining parameters of the distribution, $CV_v = \sigma_v/\mu_v$, is kept below 0.35 for all v. Consequently, the effect of truncating the original Gaussian distribution is negligible and μ_v, σ_v and CV_v are close estimates of the mean, standard deviation and coefficient of variation, respectively.
4. To satisfy the GoS requirement, the amount of capacity provisioned to serve each demand is required to be no less than its mean, i.e., $\underline{d}_v = \mu_v$.
5. All demands v are assumed to have the same mean, i.e., $\mu_v = \bar{\mu}$ for all v. Then $\sum_{v \in \mathcal{V}} \mu_v h_v = \bar{\mu} \sum_{v \in \mathcal{V}} h_v$ is an estimate of average network demand. The ratio of this quantity to the sum of all link capacities, $\rho = (\bar{\mu} \sum_{v \in \mathcal{V}} h_v) / \sum_{l \in \mathcal{L}} C_l$, is defined to be the network *load factor*. In numerical examples we vary the load factor ρ by appropriately setting the mean demand. In the sample network (Figure 1), $\sum_{l \in \mathcal{L}} C_l = 2350$ and $\sum_{v \in V} h_v = 176$, so $\mu_v = \bar{\mu} = 13.35\rho$. For given μ_v, to vary the coeffcient of variation, CV_v, we appropriately set σ_v.

In addition to the *random* demands specified above, the network also has another set of *guaranteed* demands between all network node pairs. While we assume here that the guaranteed demand is abundant, the unit price of carrying is a (small) fixed fraction, $\phi (0 < \phi < 1)$, of the unit price for carrying random demand for the same node pair. This fraction is uniform across all node pairs in our examples and its value is set at either 0.1 or 0.2. Route selection of guaranteed demands is subject to the same constraints that apply to random demands. However, from the above assumptions on guaranteed demands, specifically, prices are proportional to minimum distances (measured by hops) and volumes are unlimited, it should be clear that it is optimal to carry such demands only on minimum-distance routes.

Observe that the service provider decides on bandwidth allocation to random demands, for a potentially higher but uncertain revenue, and guaranteed demands, for a lower but secure revenue. The optimal allocation is influenced by both demand uncertainty and the service provider's risk-averseness, as will be demonstrated later.

B. Impact of Demand Uncertainty

We start with the base case and set the risk parameter $\delta = 0.5$, and the ratio of unit prices for guaranteed and random demands for the same node pair, $\phi = 0.1$. Let the coefficient of variation (CV_v) increase from the base case value of 0.1 to $0.2, 0.3$, and 0.35, and keep all other parameters unchanged. Figures 2 and 3 demonstrate vividly the revenue implications of demand uncertainty, which are consistent with Theorem 3 in the last section. In both figures, the horizonal axis corresponds to the index of random demand, which is associated with a node pair. Given various uncertainty levels, Figure 2 plots the percentage difference of mean revenue from the benchmark, which is the mean revenue obtained at zero uncertainty ($CV_v = 0$). The mean revenue for each random demand is an implication of the optimum network-wide bandwidth provisioning computed by the procedure discussed in the preceeding sections. The figure reveals the detrimental impact of demand uncertainty as the differences are always negative and higher values of CV_v always lead to larger percentage drops. The amounts of provisioned bandwidth to these demands are shown in Figure 3, which are increasing with demand uncertainty, indicating that more network capacities are diverted from serving guaranteed demands to serving random demands. This is exactly what Theorem 3 postulated, i.e., despite the fact that demands of higher uncertainty result in lower mean revenues, more bandwidth is provisioned to random demands because their expected marginal revenue is higher due to greater likelihood of reaching higher levels.

For random demands, as uncertainty increases, the utilization of provisioned bandwidth decreases. Define link usage rate as the ratio of mean carried traffic to the amount of provisioned bandwidth on a link, both calculated after aggregation over all random demands using the link. Figure 4 plots the usage rate for every link in the network at several uncertainty values. Despite the decreased efficiency of bandwidth usage, the marginal value of link capacity, characterized by the link shadow cost, increases with uncertainty, as shown in Figure 5. This fact, surprising at first, is due to the need to provision more bandwidth to random demands when uncertainty is high, which makes capacity a more constrained resource.

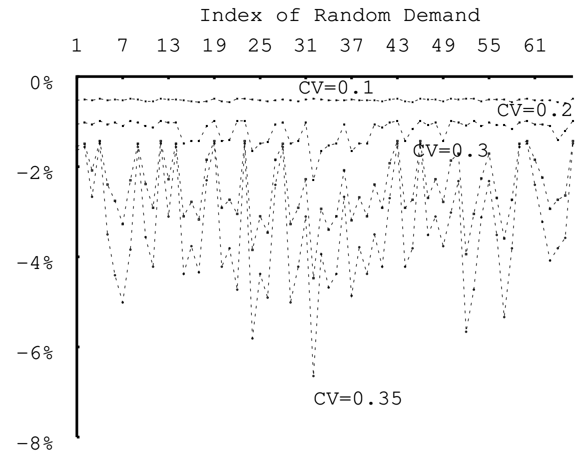

Fig. 2. Mean Revenue from Random Demand Drops with Variability

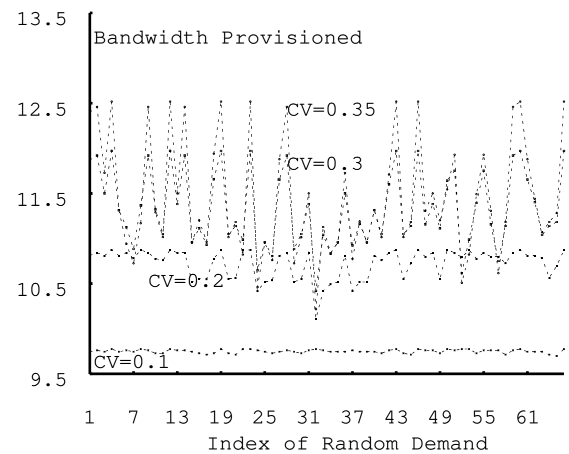

Fig. 3. Bandwidth Provisioned to Random Demand Increases with Variability

We further note that as uncertainty increases, changes of shadow costs and usage rates are notably non-uniform across all links, and depend on their location. The non-uniformity also applies to drops of mean revenue and increases of provisioned bandwidth to random demands across node pairs, i.e., links that are on admissible routes of a larger number of demands, indicating more severe competition for capacity on these links. For demands going through these links, the increase of provisioned bandwidth is smaller, and consequently, these demands have larger drops in mean revenue when the uncertainty goes up.

To summarize, increase in demand uncertainty reduces

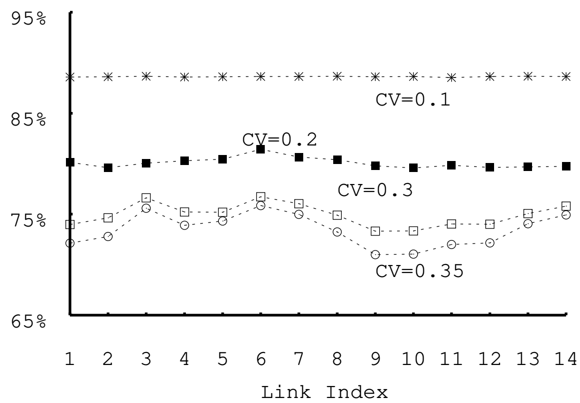

Fig. 4. Link Utilization Decreases with Variability

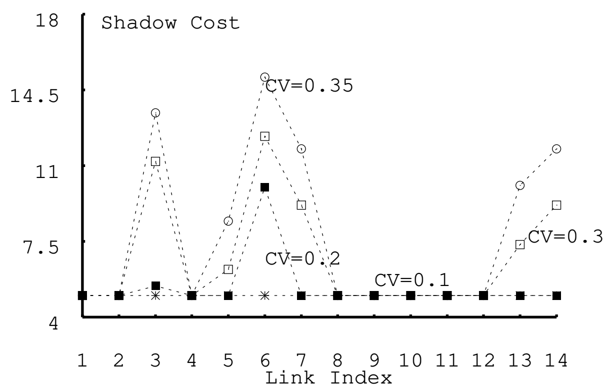

Fig. 5. Shadow Cost Increases with Variability

mean revenue from random demands and induces the service provider to allocate more capacity to these demands.

C. Risk Mitigation by Bandwidth Allocation

In this subsection, we highlight a risk mitigation mechanism derived from adjusting bandwidth allocations to the guaranteed and random demands. The former has no uncertainty, its associated revenue is risk-free, albeit lower, so allocating more bandwidth to it mitigates the revenue risk. We vary the risk-averseness parameter, δ, and investigate the implied allocation of bandwidth to the two types of demands in the optimal solution. The results show that as the service provider's attitude varies from risk-neutral to highly risk-averse, the proportion of bandwidth allocated to guaranteed demands increases, as does the proportion of revenue from guaranteed demands in the total expected revenue.

Table I gives the results for δ ranging from 0 to 2.4. For random demands, we set $\rho = 0.65$ and $CV_v = 0.35$ for all v. The ratio of prices for carrying guaranteed and random demands for the same node pair, ϕ is fixed at 0.2. In all cases, the parameter $\underline{d}_v = \mu_v$ for all v in the GoS constraint, see (12). This implies that 1531 units of capacity, out of the total of 2350, is set aside for random services to meet the GoS constraint. The service provider allocates the remaining 819 units of bandwidth to random and guaranteed demands. Table I gives *adjusted* numerical values of bandwidth and expected revenue corresponding to the excess over the respective fixed quantities committed to satisfy the GoS constraint.

TABLE I

EFFECT OF RISK AVERSENESS ON *ADJUSTED* BANDWIDTH ALLOCATION AND MEAN REVENUE COMPOSITION

δ	bandwidth allocation		mean reveue composition	
	random	guarant.	random	guarant.
0.0	49.0%	51.0%	62.5%	37.5%
0.4	46.5%	53.5%	60.6%	39.4%
0.8	43.8%	56.2%	58.6%	41.4%
1.2	40.8%	59.2%	56.2%	43.8%
1.6	37.2%	62.8%	53.2%	46.8%
2.0	35.0%	65.0%	51.2%	48.8%
2.4	32.8%	67.2%	49.1%	50.9%

The table shows that a risk-neutral carrier ($\delta = 0$) should provision 49% of uncommitted capacity to serve random demands. As the carrier becomes more risk-averse, the percentage of bandwidth allocated to random demands declines. The last row of the table shows that a conservative carrier ($\delta = 2.4$) should allocate no more than 32.8% of the uncommitted capacity to random demands. Differences in provisioning lead to differences in revenue composition. When $\delta = 0$, 62.5% of the revenue is generated from random demands. When $\delta = 2.4$, the correponding number is 49.1%.

In subsection VI-B, we showed that, for given δ, the optimal amount of bandwidth provisioned to random demands increases with demand uncertainty. We now investigate the dependence on δ. For $\delta = 0.0, 1.2, 2.4$, Figure 6 plots percentages of total link capacities provisioned to random demands, when the measure of uncertainty (CV_v) changes from 0 to 0.35. When there is no demand uncertainty ($CV_v = 0$), risk tolerance is irrelevant to the provisioning decision. As a result, the three cases with different δ values converge to the same provisioning decision that allocates about 65% of total bandwidth to random demands, and leaves the remaining 35% for guaranteed demands. When uncertainty increases, more bandwidth is shifted to random demands, but the change is significantly slower when the service provider is more risk-averse.

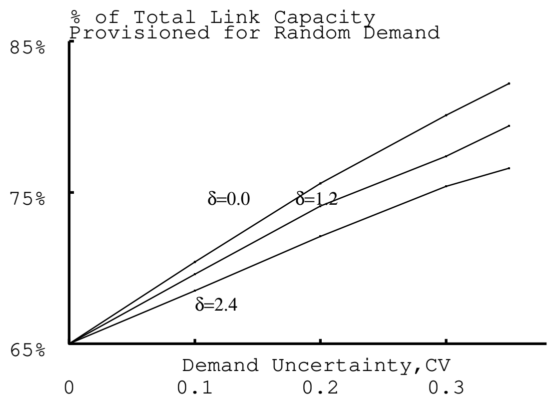

Fig. 6. Increase of Provisioned Bandwidth for Random Demands with Variability for Different Risk Tolerance

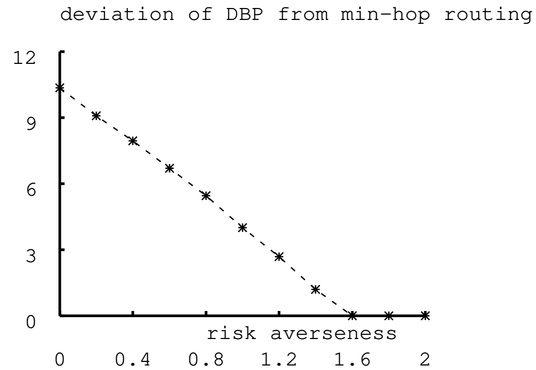

Fig. 7. Change of Route Selection with Risk-Averseness

D. Impact of Risk-Averseness on Route Selection

The preceding subsection focused on the influence of risk-averseness on bandwidth allocations. In this subsection, we illustrate that risk-averseness also influences route selection, and thereby affecting all important elements of traffic engineering design. We associate with any routing solution its *distance-bandwidth product* (DBP),

$$DBP = \sum_{v \in \mathcal{V}} \sum_{r \in \mathcal{R}(v)} \xi_r |r|,$$

where ξ_r is the amount of bandwidth provisioned on route $r \in \mathcal{R}(v)$ as defined in section II and $|r|$ is the number of hops in route r. A reference value is

$$DBP_0 = \sum_{v \in \mathcal{V}} h_v \sum_{r \in \mathcal{R}(v)} \xi_r,$$

which is the distance-bandwidth product when bandwidth is provisioned only on routes with the minimum number of hops. We let $\Delta(DBP) = (DBP - DBP_0)$ measure the departure of the routing solution from minimum-distance routing. In Figure 7, we show the dependence of $\Delta(DBP)$ on the risk-averseness parameter δ. The noteworthy feature of the figure is the strictly monotonic decrease of $\Delta(DBP)$ with increasing δ until it reaches 0, where the optimal solution coincides with the minimum-distance routing. Note that $\Delta(DBP)$ sticks to 0 for higher values of δ.

Figure 8 and Table II provide additional insight to corroborate the results in Figure 7. We consider routing of demand from node h to node e on three candidate routes, route 1 ({h-g-f-d-e}), which has the minimum distance, routes 2 ({h-g-j-k-l-e}) and 3 ({h-i-j-k-l-e}), which have an additional hop. Table II shows that with risk-neutrality ($\delta = 0$), roughly one-third of the demand is carried over

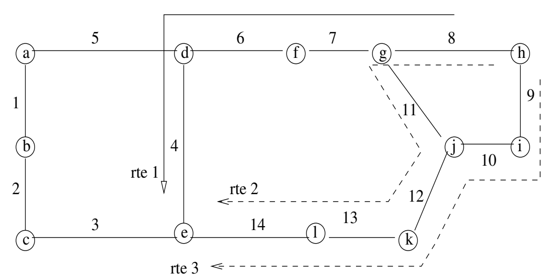

Fig. 8. Minimum Distance and Candidate Routes for Demand between (h,e)

the two non-minimum-distance routes. For $\delta \geq 1.6$, the routing is exclusively minimum distance.

While the prevalence of minimum-distance routing with risk-averseness is a phenomenon specific to our assumptions, the examples show that service provider's attitudes towards revenue risk do affect their routing decision.

E. Efficient Frontier

The effect of the service provider's tolerance to risk is summarized by the *efficient frontier*, which is obtained by solving the optimization model for various values of δ. In Figure 9 we show the frontier for the base case network. Each point on the curve gives the optimum combination of mean and standard deviation of revenue at a given value of δ, and represents the maximum expected revenue (reward) obtainable at a given level of risk, or the minimum risk the service provider has to take for a given reward.

A key feature to observe in the displayed efficient frontier in Figure 9 is the "knee" of the curve, which corresponds approaximately to mean revenue of 11000. To the left of this point, the tradeoff between risk and mean rev-

TABLE II
EFFECT OF RISK AVERSENESS ON ROUTE SELECTION

δ	min-hop	non min-hop	
	rte. 1	rte. 2	rte. 3
0.0	67.8%	5.4%	26.8%
0.4	76.1%	7.2%	16.7%
0.8	86.4%	6.4%	7.2%
1.2	93.0%	3.5%	3.5%
1.6	100.0%	0.0%	0.0%

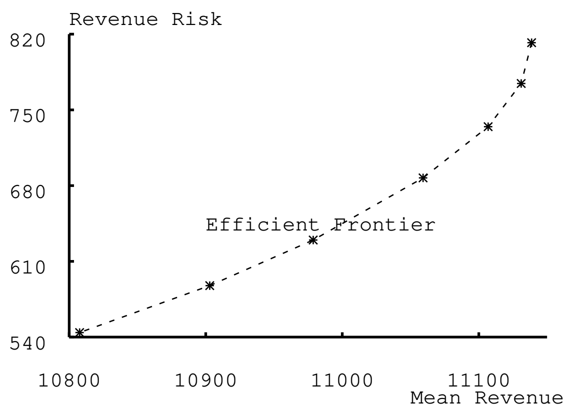

Fig. 9. Efficient Frontier

enue is roughly linear and benign, and to the right the curve is increasingly nonlinear and any small gain in the expected revenue is accompanied by a subtantial cost in increased risk. Moreover, the figure highlights the importance of precision in locating the knee since a relatively small variation of about $1\% - 2\%$ in mean revenue around the knee can lead to subtantial changes in the slope of the efficient frontier.

VII. CONCLUSION

We have presented and analyzed a stochastic traffic engineering framework for off-line planning of bandwidth provisioning and routing. The framework is based on an optimization model that uses probability distributions of demands as inputs and maximizes the weighted combination of the mean revenue and the risk of revenue shortfall. We discuss properties of the objective function, and strategies for solving the model as a concave maximization problem. In our numerical studies, we analyze the impacts of demand uncertainty on various aspects of traffic engineering design. We observe significant changes in mean revenue, bandwidth provisioning, link shadow costs and utilization with demand uncertainty. We further demonstrate that changes in bandwidth provisioning with uncertainty are strongly influenced by the service provider's tolerance to risk, and give the efficient frontier, which characterizes the tradeoff between expected revenue and revenue risk.

Our analysis can be extended in several directions. For instance, an extension of the model can give service providers an optimization tool for use in interactions with peers in bandwidth sharing and network interconnections. These extensions are currently being investigated, and will be reported in future publications.

REFERENCES

[1] V. Agrawal and S. Seshadri, Impact of Uncertainty and Risk Aversion on Price and Quantity in the Newsvendor Problem, *M & SOM*, vol. 2, no. 4, pp. 410-423, Fall 2000.

[2] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows.* Prentice-Hall, Inc., Upper Saddle River, NJ, 1993.

[3] D. Applegate and M. Thorup, Load Optimal MPLS Routing with N+M Labels, *Proceedings of IEEE INFOCOM2003*, San Francisco, April 2003.

[4] P. Aukia, M. Kodialam, P. V. N. Koppol, T. V. Lakshman, H. Sarin, and B. Suter, RATES: A Server for MPLS Traffic Engineering, *IEEE Network*, **14**, pp. 34-41, March/April 2000.

[5] D. O. Awduche, A. Chiu, A. Elwalid, I. Widjaja, and X. Xiao, Overview and Principles of Internet Traffic Engineering *RFC 3272, IETF*, May 2002.

[6] J. Cao, D. Davis, S. V. Wiel, and B. Yu, Time-Varying Network Tomography, *Journal of the American Statistical Association*, **95**, pp. 1063-1075.

[7] J. Cao, S. V. Wiel, B. Yu, and Z. Zhu, A Scalable Method for Estimating Network Traffic Matrices, *Bell Labs Technical Memorandum*, 2000.

[8] A. Elwalid, C. Jin, S. Low, and I. Widjaja, Mate: MPLS Traffic Engineering, *Proceedings of IEEE INFOCOM2001*, pp. 1300-1309, 2001.

[9] A. Elwalid, D. Mitra, I. Saniee and I. Widjaja, Routing and Protection in GMPLS Networks: From Shortest Paths to Optimized Designs, *Journal of Lightwave Technology*, vol. 21, no. 11, pp. 2828-2838, Nov. 2003.

[10] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, and J. Rexford, NetScope: Traffic Engineering for IP Networks, *IEEE Network*, **14**, pp. 11-19, March/April 2001.

[11] B. Fortz and M. Thorup, Internet Traffic Engineering by Optimizing OSPF Weights, *Proceedings of IEEE INFOCOM2000*, pp. 519-528, 2000.

[12] M. Grossglauser and J. Rexford, Passive Traffic Measurement for IP Operations *INFORMS Telecom Meeting 2002*, Ft. Lauderdale, FL, March, 2002.

[13] J. Hirshleifer and J. G. Riley, *The Analytics of Uncertainty and Information*, Cambridge University Press, Cambridge, England, 1992.

[14] F. Kelly, A. Maulloo, and D. Tan, Rate Control in Communication Networks: Shadow Prices, Proportional Fairness and Stability, *Journal of th e Operational Research Society*, 49, pp. 237-252, 1998

[15] C. Lagoa and H. Che, Decentralized Optimal Traffic Engineering in the Internet, *Computer Communications Review*, vol. 30, No. 5, pp. 39-47, October 2000.

[16] H. Levy, *Stochastic Dominance: Investment Decision Making Under Uncertainty*, Kluwer Academic Publishers, MA, 1998.

[17] Level 3 website, Level 3 Signs New Agreement with Microsoft, *www.level3.com/press/2053.html*, June 2001.

[18] Level 3 website, (3)Crossroads Wholesale Internet Access, *www.level3.com/560.html*.

[19] S. G. Lanning, D. Mitra, Q. Wang and M. H. Wright, Optimal Planning for Optical Transport Networks, *Philosophical Transactions of The Royal Society*, London, 2000, pp. 2183-2196.

[20] D. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, Reading, MA, 1973.

[21] H. Markowitz, *Mean Variance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell, New York, 1987.

[22] R. Morris and D. Lin, Variance of Aggregated Web Traffic, *Proceedings of IEEE INFOCOM2000*, Tel Aviv, April 2000.

[23] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic Matrices Estimation: Existing Techniques and New Directions", *Proceedings of ACM SIGCOMM* Pittsburgh. August 2002.

[24] 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.

[25] 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.

[26] D. Mitra, K. G. Ramakrishnan, and Q. Wang, Combined Economic Modeling and Traffic Engineering: Joint Optimization of Pricing and Routing in Multi-Service Networks, *Traffic Engineering in the Internet Era: Proceedings of 17th International Traffic Congress-ITC17*, eds. J.M. de Souza, N.L.S. da Fonseca and E.A. de Souza e Silva, Elsevier, pp. 73-85, 2001.

[27] D. Mitra and Q. Wang, Stochastic Traffic Engineering, with Applications to Network Revenue Management, *Proceedings of IEEE INFOCOM03*, San Francisco, April 2003.

[28] D. Mitra and Q. Wang, Risk-Aware Network Profit Management in a Two-Tier Market, *Proceedings of 18th International Telegraffic Congress - ITC18*, Berlin, August-September 2003.

[29] W. Ogryczak and A. Ruszczynski, Dual stochastic dominance and related mean–risk models, *SIAM Journal on Optimization*, vol. 13 pp. 60–78, 2002.

[30] N. Semret, R. R.-F. Liao, A. T. Campell, and A. A. Lazar, Peering and Provisioning of Differentiated Internet Services, *Proceedings of IEEE INFOCOM 2000*, Tel Aviv, pp. 100-107, April 2000.

[31] S. Suri, M. Waldvogel, and P. R. Warkhede, Profile-Based Routing: A New Framework for MPLS Traffic Engineering, *QofIS'01*, Coimbra, Portugal, 2001.

[32] P. Trimintzios, I. Andrikopoulos, G. Pavlou, P. Flegkas, D. Griffin, P. Georgatsos, D. Goderis, Y. T'Joens, L. Georgiadis, C. Jacquenet, and R. Egan, A Management and Control Architecture for Providing IP Differentiated Services in MPLS-Based Networks, *IEEE Communications Magazine*, vol. 39, no. 5, pp. 80-87, May 2001.

[33] 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, 1996.

[34] J. von Neuman and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, New Jersey, 1953.

[35] W. Wang, M. Palaniswami and S. H. Low, Optimal Flow Control and Routing in Multi-path Networks, *Performance Evaluation*, vol. 52, No. 2-3, pp. 119-132, April 2003.