Computer Communications 31 (2008) 3169–3177

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Optimization driven bandwidth provisioning in service overlay networks

Kyung-Joon Park [a,*], Chong-Ho Choi [b,1]

[a] Department of Computer Science, University of Illinois, 201 N. Goodwin Avenue, Urbana, IL 61801-2302, USA
[b] School of Electrical Engineering and Computer Science, Seoul National University, Seoul 151-742, Republic of Korea

ARTICLE INFO

Article history:
Received 10 August 2007
Received in revised form 21 April 2008
Accepted 30 April 2008
Available online 15 May 2008

Keywords:
Service overlay networks
Bandwidth provisioning
Convex optimization

ABSTRACT

We consider the bandwidth provisioning problem in service overlay networks (SON). We formulate the bandwidth provisioning problem in an optimization framework taking the stochastic processes involved into consideration. First, we provide the optimal solution that maximizes the expected net revenue in a single-link network under stationary traffic demand. Then, we show that the revenue maximization problem in a general topology under stationary traffic can be nicely approximated by a *separable convex* optimization problem. We derive an approximate optimal solution for a general topology and provide a bandwidth provisioning algorithm via a gradient method, which works in a *distributed* manner. We also verify the effectiveness of the approximate optimal solution by showing several important characteristics of the solution. The performance of the algorithm is investigated through extensive simulations.

1. Introduction

Recently, the notion of service overlay networks (SON) has been widely applied to various areas of the Internet. *Overcast* [1] adopted an overlay structure for the application-level multicasting service. *Scattercast* [2] is a broadcasting infrastructure that forms an overlay network. *Detour* [3] and *Resilient Overlay Networks* (RON) [4] employ overlay techniques to improve routing support. Overlay network architectures have also been employed in peer-to-peer services [5]. In addition, overlay techniques have attracted much attention from industries [6,7] as a method to provide end-to-end Quality-of-Services (QoS) over the current Internet. In fact, the concept of overlay networks is quite old and the Internet itself was devised as an overlay network on the telephone network. The SON structure considered here is a logical generalization of these architectures.

A SON consists of connections between *service gateways* that forward packets and perform control functions. The logical connection between two service gateways is offered by the underlying network domain, characterized by a certain amount of bandwidth and other QoS constraints. QoS constraints are specified in a Service Level Agreement (SLA) between a SON and the network domain. Based on the SLA, a SON can guarantee the end-to-end QoS via appropriate bandwidth provisioning for each logical link. In the deployment of a SON, it is a critical issue for the SON provider to maximize the net revenue. Among the costs incurred in the deployment of a SON, a major one is the cost of purchasing bandwidth from the network domains. Meanwhile, a SON must provide an appropriate amount of bandwidth to satisfy the end-to-end QoS constraints. Hence, there is a tradeoff between the bandwidth cost and the QoS guarantee in the revenue-maximizing problem of a SON. Consequently, it is crucial to find a way to resolve this bandwidth provisioning problem in the deployment of a SON.

The bandwidth provisioning problem has been studied extensively in various frameworks [8–17]. In this paper, we consider the bandwidth provisioning problem in a SON as the work in [8,9], where the bandwidth provisioning problem was considered as an effective means to provide end-to-end QoS services and has been successfully formulated and approximately solved within an optimization framework. However, the necessary condition in [8,9], which constitutes a set of coupled equations, is difficult to solve as the number of links increases. Furthermore, there is not enough information on the problem structure to guarantee the convergence of the algorithm in [8,9]. In this work, we rigorously formulate the bandwidth provisioning problem in an optimization framework with stochastic processes involved. Then, we obtain a necessary and sufficient optimality condition in a single-link topology with the wide-sense stationary assumption of the traffic demand. Then, we incorporate a gradient algorithm for obtaining the optimal solution. Furthermore, we show that the problem in a general topology can be approximately constructed as a *separable convex* optimization problem. Consequently, as in the single-link case, we develop a bandwidth provisioning algorithm in a general topology, which works in a *distributed* manner.

The main contribution of this paper is as follows: First of all, we rigorously formulate the SON bandwidth provisioning problem in an optimization framework and construct an approximate

* Corresponding author. Tel.: +1 217 419 6556; fax: +1 217 244 6500.
E-mail addresses: kjp@uiuc.edu, kyungjoon.park@gmail.com (K.-J. Park), chchoi@csl.snu.ac.kr (C.-H. Choi).
[1] Tel.: +82 2 880 7313; fax: +82 2 885 4459.

0140-3664/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2008.04.028

optimization problem that is shown to be *separable* and *convex*. From the convexity, we can guarantee the convergence of the gradient algorithm. Also, the curse of dimensionality problem, which is an important issue for scalability, can be circumvented by the separability of the problem. Furthermore, we derive important mathematical relations between the original and the approximate problems, which show the effectiveness of the approximate optimal solution.

The remaining part of the paper is organized as follows: In Section 2, we give a detailed explanation on the bandwidth provisioning problem in an optimization framework and introduce the notations and the net revenue function. Then, we propose the optimal solution in a single-link topology and formulate an approximate optimization problem in a general topology, which is shown to be separable and convex. We also show several important characteristics of the approximate optimal solution. Extensive simulation results are given in Section 3. Finally, conclusions follow in Section 4.

2. Bandwidth provisioning problem in SON

In this section, we formulate the bandwidth provisioning problem within an optimization framework taking into consideration the QoS violation and the network revenue. Note that the formulation of the net revenue in a stochastic framework is more rigorous than those in [8,9]. Based on the formulation, we show that the revenue maximization problem can be nicely approximated by a separable convex optimization problem. Hence, we can guarantee the convergence of a gradient algorithm that finds an optimal solution for the approximate problem, which is just slightly larger than the optimal solution for the original problem.

2.1. A logical model of SON

With the pipe model, a SON provides the required amount of bandwidth between any two service gateways across a network domain, i.e., a pipe with a certain amount of bandwidth is provided between any two service gateways. Since our formulation of the bandwidth provisioning problem is based on the work in [8,9], we use the same logical model of a SON. We represent the *logical* path from a service gateway u to an adjacent service gateway v across a network domain D by a three-tuple $l = (u,v,D)$ and consider l as a *link* between u and v across D. We assume that a certain amount of bandwidth is assigned to each path between a SON and the access networks which act as sources and sinks of traffic. We can consider each access network A as a *virtual* service gateway u_A. Then, as in the service gateway case, we can simply represent the connection between u_A and an adjacent gateway v across A by a corresponding link (u_A,v,A). For a given logical link $l = (u,v,D)$, a SON provider will make a contract with the underlying network domain D guaranteeing a certain amount of bandwidth c_l. The bandwidth provisioning problem of the SON would then be to decide the bandwidth of each link $l = (u,v,D)$ so that its overall revenue can be maximized with a certain level of end-to-end QoS guarantee.

Although the QoS requirement for a SON service can be assessed by quite diverse metrics such as loss probability, delay or delay jitter, control of link utilization is a key issue in providing QoS in most cases. In other words, it is required to ensure that the overall load on a link does not exceed a certain level of pre-specified link utilization. Consequently, we simply consider the problem of determining the level of link utilization for QoS guarantee of a SON.

With the level of link utilization as a QoS metric, there is a tradeoff between seeking the maximum revenue and guaranteeing the QoS simultaneously. If we increase the link bandwidth c_l, then, we can guarantee a high level of QoS. Yet, at the same time, the revenue will decrease as a result of the additional cost for purchasing bandwidth, and vice versa. Hence, the bandwidth provisioning problem now becomes an optimization problem of finding an optimal link bandwidth that maximizes the revenue with QoS constraints. This will be explained in detail in the next section.

2.2. Bandwidth demand, QoS violation, and net revenue

Now, we rigorously formulate the bandwidth provisioning problem in an optimization framework, taking into consideration the stochastic processes involved. To this end, we introduce several notations. For a set A, let $|A|$ denote its cardinality. Consider a set L of links, indexed by $l = 1,2, \dots ,|L|$ with capacities c_l. Also think of a set S of traffic sources, indexed by $s = 1,2, \dots ,|S|$. For each source s, let X_s denote the traffic demand of source s averaged over the n-th micro-time slot. The micro-time unit is denoted by Δ_m. Note that the micro-time unit Δ_m is relatively very small compared to the macro-time unit, denoted by Δ_M. Hereafter, the period Δ_m is considered as the basic time unit for traffic demand and Δ_M for bandwidth provisioning, respectively. Furthermore, for each l, let Y_l denote the traffic demand for link l, then $Y_l = \sum_{s \in S(l)} X_s$ where $S(l) \subseteq S$ is a set of s which traverses link l. Here, let $f_X(x)$ denote the probability density function (pdf) of X.

We assume that a SON receives g_s amounts of money from source s per unit of traffic per unit time. On the other hand, a SON pays $\Phi_l(c_l)$ per unit time for reserving c_l amount of bandwidth for link l. We refer to $\Phi_l(c_l)$ as the bandwidth cost function of link l. Further, for each source s, let π_s denote the penalty per unit of traffic demand per unit time when the QoS is violated.

Now, our main concern is to maximize the expected net revenue by adjusting the values of c_l's, $l = 1,2, \cdots ,|L|$. With all these notations, we can express the net revenue of a SON as follows:

$$W = \sum_s g_s X_s - \sum_l \Phi_l(c_l) - \sum_s \pi_s X_s I_s,$$

where I_s is the penalty indicator function of source s with

$$I_s = \begin{cases} 1 & \text{if} \quad \exists l \in L(s) \quad \text{such that} \quad Y_l > c_l, \\ 0 & \text{otherwise}, \end{cases}$$

and $L(s) \subseteq L$ is a set of l which are traversed by source s. The objective is to choose link bandwidths bi c $= (c_l, l \in L)$ such that

$$\begin{aligned} &\text{maximize} \quad \mathbb{E}[W] \\ &\text{subject to} \quad c_l \geqslant m_{Y_l}, \quad \forall l \in L, \end{aligned} \tag{1}$$

where $m_{Y_l} = \mathbb{E}[Y_l]$. The constraint in (1) suggests that the average aggregate traffic demand at any link should not exceed the bandwidth. In the following section, we derive an algorithm for obtaining the optimal solution in a single-link topology and develop a distributed algorithm for an approximate optimal solution in a general topology.

2.3. Single-link topology: optimal solution

In a single-link topology, we have

$$W = g_1 X_1 - \Phi_1(c_1) - \pi_1 Z_1, \tag{2}$$

where $Z_1 = X_1 I_1$. For notational simplicity, we omit subscripts hereafter. By taking expectation in (2),

$$\mathbb{E}[W] = gm - (\Phi(c) + \pi\mathbb{E}[Z]), = gm - V(c),$$

where $m = \mathbb{E}[X]$, and $V(c) := \Phi(c) + \pi \int_c^\infty x f_X(x) dx$. Hence, we have

$$\begin{aligned} &\text{maximize} \quad \mathbb{E}[W] = gm - V(c) \\ &\text{subject to} \quad c \geqslant m. \end{aligned} \tag{3}$$

Consequently, instead of maximizing $\mathbb{E}[W]$, now we can solve the optimization problem by minimizing $V(c)$ in (3) as follows:

$$\begin{aligned}&\text{minimize } V(c)\\&\text{subject to } c \geqslant m.\end{aligned}$$

The necessary optimality condition for minimizing $V(c)$ is

$$\frac{\mathrm{d}V(c)}{\mathrm{d}c} = \frac{\mathrm{d}\Phi(c)}{\mathrm{d}c} + \pi\frac{\mathrm{d}}{\mathrm{d}c}\int_c^\infty x f_X(x)\mathrm{d}x = 0.$$

From the Leibniz integral rule, we have

$$\frac{\mathrm{d}V(c)}{\mathrm{d}c} = \frac{\mathrm{d}\Phi(c)}{\mathrm{d}c} - \pi c f_X(c) = 0. \tag{4}$$

Now, assume that $\Phi(c)$ is linear, i.e., $\Phi(x) = \phi x$ as in [9] and $f_X(x)$ is Gaussian. Gaussian assumption of $f_X(x)$ is reasonable since X is aggregate traffic with high degree of multiplexing. With these assumptions, we can express the necessary optimality condition (4) as

$$\kappa - c f_X(c) = 0, \tag{5}$$

where $\kappa = \phi/\pi$ and $f_X(c) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{(c-m)^2}{2\sigma^2}]$.

There are three cases to the solutions to (5) according to the value of κ. The first case is when (5) has no solution and $\frac{\mathrm{d}V(c)}{\mathrm{d}c} > 0$, $\forall c \in \Re_+$, which happens with a too large value of κ. Similarly, the second case is when (5) has only one solution and $\frac{\mathrm{d}V(c)}{\mathrm{d}c} \geqslant 0$, $\forall c \in \Re_+$. Hence, for these two cases, $V(c)$ attains its minimum at $c = 0$, under which bandwidth provisioning has no practical meaning. Only when κ is in a reasonable regime, in which there are two real solutions to (5), the overall problem of bandwidth provisioning is practically meaningful. This case is illustrated in Fig. 1, where the mean and the variance of the traffic demand X is given as $m = 5$ and $\sigma^2 = 1$, respectively. Furthermore, $\kappa = 1$ and the two solutions to (5) is denoted by $c^{*,1}$ and $c^{*,2}$.

Now, we show that $c^{*,2}$ in Fig. 1 is the unique minimizing solution of $V(c)$. First, we obtain the value of $\underline{c}$ in Fig. 1 by solving

$$\frac{\mathrm{d}^2V}{\mathrm{d}c^2} = -\left[f_X(c) + c\frac{\mathrm{d}f_X(c)}{dc}\right] = 0.$$

Since $\mathrm{d}f_X(c)/dc = -(c - m/\sigma^2)f_X(c)$, $\underline{c}$ is the solution of $c(c - m) = \sigma^2$. Hence, we have

$$\underline{c} = \frac{m + \sqrt{m^2 + 4\sigma^2}}{2}.$$

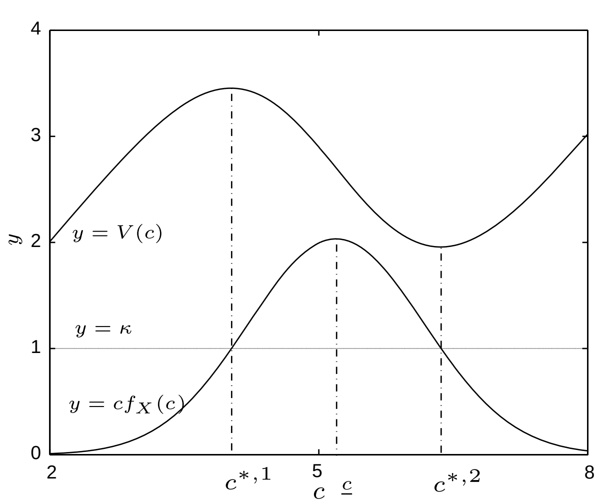


Fig. 1. Relation between the solutions $c^{*,1}$ and $c^{*,2}$ of $\mathrm{d}V(c)/\mathrm{d}c = 0$ and the pricing parameter κ.

Note that the overall bandwidth provisioning problem is practically meaningful only when $\kappa < \overline{\kappa}$ where

$$\overline{\kappa} = \underline{c}f_X(\underline{c}) = \frac{1}{2\sqrt{2\pi}}\left(\frac{1+\sqrt{4\delta^2+1}}{\delta}\right)\exp\left[-\frac{1}{2}\left(\frac{\sqrt{4\delta^2+1}-1}{\delta}\right)^2\right],$$

and $\delta = \sigma/m$. Consequently, we have the following proposition:

Proposition 1. *A solution to (5) exists if $\kappa \leqslant \overline{\kappa}$.*

Remark 1. Proposition 1 provides a useful guideline on whether bandwidth provisioning is necessary or not. A SON does not need to purchase any bandwidth if $\kappa > \overline{\kappa}$. This indicates that the penalty from link over-utilization is trivial compared to the bandwidth cost. In practice, Proposition 1 may be a helpful guideline for a SON provider to decide whether over-provisioning is necessary or not.

From Fig. 1, we know that $c^{*,1} < \underline{c} < c^{*,2}$. If $\sigma/m \ll 1$, then $\underline{c} \approx m$ and consequently $c^{*,1} < m < c^{*,2}$. Accordingly, from Fig. 1, we know that $c^{*,2}$ is the unique optimal solution in $[m,\infty]$. We can conclude that $c^{*,2}$ is the unique global solution c^* for maximizing $\mathbb{E}[W]$ in the practically meaningful parameter region of κ, i.e., $\kappa \leqslant \overline{\kappa}$. An algorithm to find c^* by a gradient method is given in Algorithm 1, in which γ is the step size and ϵ is a pre-specified small positive number.

2.4. General topology: approximate optimal solution

Now, consider a solution for a general topology. First, let $I_l(Y_l)$ be an indicator function for the link l with

$$I_l(Y_l) = \begin{cases}1 & \text{if } \quad Y_l > c_l,\\ 0 & \text{otherwise}.\end{cases}$$

Then, we have

$$I_s = \bigvee_{l \in L(s)} I_l(Y_l),$$

where $\bigvee_{l\in L(s)} I_l := I_1 \vee I_2 \vee \cdots \vee I_{|L(s)|}$ and $\vee$ denotes the logical OR operation. Now,

$$\begin{aligned}W &= \sum_s g_s X_s - \sum_l \Phi_l(c_l) - \sum_s \pi_s X_s I_s,\\ &= \sum_s g_s X_s - \left[\sum_l \Phi_l(c_l) + \sum_s \pi_s X_s \bigvee_{l\in L(s)} I_l(Y_l)\right].\end{aligned}$$

The optimization problem becomes

$$\begin{aligned}&\text{maximize } \mathbb{E}[W] = \sum_s g_s \overline{X}_s - V(\mathbf{c})\\&\text{subject to } c_l \geqslant m_{Y_l}, \quad \forall l \in L,\end{aligned}$$

where $V(\mathbf{c}) := \mathbb{E}\left[\sum_l \Phi_l(c_l) + \sum_s \pi_s X_s \left(\bigvee_{l\in L(s)} I_l(Y_l)\right)\right]$, $m_{Y_l} = \mathbb{E}[Y_l] = \sum_{s\in S(l)} m_{X_s}$, and $m_{X_s} = \mathbb{E}[X_s]$. Hence, the maximization problem becomes

$$\begin{aligned}&\text{minimize } V(\mathbf{c}) = \sum_l \Phi_l(c_l) + \sum_s \pi_s \mathbb{E}\left[X_s \bigvee_{l\in L(s)} I_l(Y_l)\right]\\&\text{subject to } c_l \geqslant m_{Y_l}, \forall l \in L.\end{aligned} \tag{6}$$

Algorithm 1: Gradient algorithm to find c^*

$\hat{c}_0 \leftarrow \frac{m+\sqrt{m^2+4\sigma^2}}{2}$
while $|\hat{c}_{n+1} - \hat{c}_n| > \epsilon$ **do**
  $\hat{c}_{n+1} \leftarrow \hat{c}_n - \gamma(\kappa - \hat{c}_n f_X(\hat{c}_n))$
end while

We restrict our attention to the region of $[c_l^m, c_l^M]$ for each link's bandwidth c_l where $c_l^m = m_{Y_l}$ and c_l^M is a sufficiently large number, then the domain of $V(\mathbf{c})$ becomes $A = [c_1^m, c_1^M] \times \cdots \times [c_{|L|}^m, c_{|L|}^M] \subset \Re_+^{|L|}$. We have the following result for the existence of the solution to (6):

Proposition 2. *If $\Phi_l(c_l)$ is an increasing function in $\Re_+$ and $\lim_{c_l \to \infty} \Phi_l(c_l) = \infty$ for $\forall l \in L$, there exists a vector $\mathbf{c}^* \in A$ such that $V(\mathbf{c}^*) = \inf_{\mathbf{c} \in A} V(\mathbf{c})$. Furthermore, there exists a vector $\mathbf{c}^* \in A$ such that $\nabla V(\mathbf{c}^*) = 0$.*

Proof. Since A is closed, we show that V is coercive. Let a sequence $\{\mathbf{c}^k\}$ denote an element of A such that $\|\mathbf{c}^k\| \to \infty$ as $k \to \infty$. Then, for $\mathbf{c}^k = (c_1^k, \cdots, c_{|L|}^k)$, there exists $i \in L$ such that $c_i^k \to \infty$. Hence, for any given large number $K_1 \in \Re_+$, we can always find $j \in L$ such that $\Phi_j(c_j^k) > K_1$ if $k \geqslant N(K_1)$. Since $V(\mathbf{c}^k) \geqslant \Phi_i(c_i^k)$ always holds for $\forall i \in L$, for any given $K_1 \in \Re_+$, we have $V(\mathbf{c}^k) > K_1$ if $k \geqslant N(K_1)$, i.e., $\lim_{k\to\infty} V(\mathbf{c}^k) = \infty$. Hence, V is coercive. From the Weierstrass theorem [18], there exists a vector $\mathbf{c}^* \in A$ such that $V(\mathbf{c}^*) = \inf_{\mathbf{c} \in A} V(\mathbf{c})$. Furthermore, Since V is continuously differentiable in the compact set A, $\mathbf{c}^* = \arg\inf_{\mathbf{c} \in A} V(\mathbf{c})$ in Proposition 2 satisfies $\nabla V(\mathbf{c}^*) = 0$ as the first-order necessary optimality condition [18]. □

Even though Theorem 2 guarantees the existence of the solution to (6), it is a difficult task to find one in practice. The difficulty is due to the fact that the bandwidth cost $\Phi_l(c_l)$ is considered on the basis of a link while the penalty $X_s I_s$ is considered on the basis of a source. Hence, we need to transform the penalty of each source into that of each link. To this end, we introduce the following approximation $\hat{I}_s$ for I_s:

$$\widehat{I}_s = \sum_{l \in L(s)} I_l(Y_l). \tag{7}$$

Note that we will verify the effectiveness of (7) in Theorem 4 and 5. Now, we approximate $V(\mathbf{c})$ with $\hat{V}(\mathbf{c})$ by using (7) as follows:

$$\begin{aligned}\hat{V}(\mathbf{c}) &= \sum_l \Phi_l(c_l) + \mathbb{E}\left[\sum_s \pi_s X_s \left(\sum_{l \in L(s)} I_l(Y_l)\right)\right],\\ &= \sum_l \Phi_l(c_l) + \mathbb{E}\left[\sum_l \sum_{s \in S(l)} \pi_s X_s I_l(Y_l)\right],\\ &= \sum_l \left(\Phi_l(c_l) + \sum_{s \in S(l)} \pi_s \mathbb{E}[X_s I_l(Y_l)]\right).\end{aligned} \tag{8}$$

Here, $\Phi_l(c_l)$ and $I_l(Y_l)$ depend only on c_l and (8) becomes

$$\widehat{V}(\mathbf{c}) = \sum_l \widehat{V}_l(c_l),$$

where $\widehat{V}_l(c_l) = \Phi_l(c_l) + \sum_{s \in S(l)} \pi_s \mathbb{E}[X_s I_l(Y_l)]$. Consequently, we have

$$\begin{aligned}&\text{minimize} \quad \widehat{V}(\mathbf{c}) = \sum_l \widehat{V}_l(c_l)\\ &\text{subject to} \quad c_l \geqslant m_{Y_l}, \quad \forall l \in L.\end{aligned} \tag{9}$$

Since $\widehat{V}_l(c_l)$ is a function of c_l only, the problem is *separable*.With the same assumptions on $\Phi_l(c_l)$ and $f_{X_s}(x_s)$ as in the single-link topology, the necessary optimality condition for minimizing $\widehat{V}_l(c_l)$ can be obtained as

$$\begin{aligned}\frac{d\widehat{V}_l(c_l)}{dc_l} &= \phi_l - \left[\pi_s c_l f_{Y_l}(c_l) + \sum_{j \in S(l), j \neq s} (\pi_j - \pi_s) \times \left(c_l f_{X_j}(c_l) * g_{S(l)\setminus\{j\}}(c_l)\right)\right]\\ &= 0\end{aligned} \tag{10}$$

for any $s \in S$, where $g_{S(l)\setminus\{j\}}$ is the convolution of all the f_{X_l}'s, $l \in S(l)\setminus\{j\}$. Note that $g_{S(l)\setminus\{j\}}$ becomes the Gaussian distribution when f_{X_s}'s are Gaussian. Now, we show that $\widehat{V}_l(c_l)$ is convex in a some suitable sub-domain $D \subset A$.

Lemma 1. *For two Gaussian pdf's $f(x)$ and $g(x)$ with means and variances, (m_1, σ_1^2) and (m_2, σ_2^2), respectively, let $h(x) := xf(x)^* g(x)$. Then, $\frac{dh(x)}{dx} < 0$ in $(\alpha, \infty]$ where $\alpha = \frac{1}{2}\left[(1-\gamma^2)m_1 + 2m_2 + \sqrt{((1+\gamma^2)m_1)^2 + 4(\sigma_1^2 + \sigma_2^2)}\right]$ and $\gamma = (\frac{\sigma_2}{\sigma_1})^2$.*

Proof. From $\frac{df(x)}{dx} = -\frac{(x-m_1)}{\sigma_1^2} f(x)$, we have

$$xf(x) = -\sigma_1^2 \frac{df(x)}{dx} + m_1 f(x).$$

Hence,

$$\begin{aligned}xf(x) * g(x) &= \left(-\sigma_1^2 \frac{df(x)}{dx} + m_1 f(x)\right) * g(x),\\ &= -\sigma_1^2 \frac{df(x)}{dx} * g(x) + m_1 f(x) * g(x),\\ &= -\sigma_1^2 \frac{d(f(x) * g(x))}{dx} + m_1 f(x) * g(x).\end{aligned}$$

By using the relationship $d(f(x)^*g(x))/dx = -(x-m)/\sigma^2(f(x)^*g(x))$ where $m = m_1 + m_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$,

$$\begin{aligned}xf(x) * g(x) &= (x-m)\left(\frac{\sigma_1}{\sigma}\right)^2 (f(x) * g(x)) + m_1 f(x) * g(x),\\ &= \left[(x-m)\left(\frac{\sigma_1}{\sigma}\right)^2 + m_1\right](f(x) * g(x)),\\ &= \frac{1}{\sigma^2}(\sigma_1^2 x + \sigma_2^2 m_1 - \sigma_1^2 m_2)(f(x) * g(x)).\end{aligned}$$

From $h(x) = xf(x)^*g(x)$ and $F(x) := f(x)^*g(x)$,

$$\begin{aligned}\frac{dh(x)}{dx} &= \frac{1}{\sigma^2}\left[\sigma_1^2 F(x) + (\sigma_1^2 x + \sigma_2^2 m_1 - \sigma_1^2 m_2)\frac{dF(x)}{dx}\right],\\ &= \frac{1}{\sigma^2}\left[\sigma_1^2 x^2 - (\sigma_1^2 m + \sigma_1^2 m_2 - \sigma_2^2 m_1)x\right.\\ &\quad \left. -(s_2^2 m_1 - s_1^2 m_2)m + s_1^2 s^2\right]F(x),\end{aligned} \tag{11}$$

by using $dF(x)/dx = -\{(x-m)/\sigma^2\}F(x)$. After some algebraic manipulation using (11), we can show that $dh(x)/dx = 0$ have real solutions if

$$\max\left((\sigma_1^2 m_2 - \sigma_2^2 m_1), 0\right) \leqslant \frac{1}{2}\left[\sigma_1^2 m + \frac{(\sigma_2^2 m_1 - \sigma_1^2 m_2)^2}{\sigma_1^2 m} + \frac{4\sigma_1^2 \sigma^2}{m}\right]. \tag{12}$$

Meanwhile, by the arithmetic–geometric mean inequality, the following inequality always holds:

$$\begin{aligned}[\text{RHS of}(12)] &\geqslant \frac{1}{2}\left[2|\sigma_2^2 m_1 - \sigma_1^2 m_2| + \frac{4\sigma_1^2\sigma^2}{m}\right],\\ &\geqslant |\sigma_2^2 m_1 - \sigma_1^2 m_2| \geqslant [\text{LHS of (12)}],\end{aligned} \tag{13}$$

where RHS and LHS are the right hand side and the left hand side, respectively. From (12) and (13), (11) gives

$$\frac{dh(x)}{dx} < 0 \quad \text{in} \quad (\alpha, \infty],$$

where

$$\alpha = \frac{1}{2}\left[(1-\gamma^2)m_1 + 2m_2 + \sqrt{((1+\gamma^2)m_1)^2 + 4(\sigma_1^2 + \sigma_2^2)}\right] \quad \text{and} \quad \gamma = \left(\frac{\sigma_2}{\sigma_1}\right)^2. \qquad □$$

Now, we show the convexity of $\widehat{V}_l(c_l)$.

Proposition 3. *Let us denote* $\alpha_{l,j} = \frac{1}{2}\left[(1-\gamma^2)m_j + 2m_{l,j} + \sqrt{\left((1+\gamma^2)m_j\right)^2 + 4\left(\sigma_j^2+\sigma_{l,j}^2\right)}\right]$ *where* $\gamma_{l,j} = \left(\frac{\sigma_{l,j}}{\sigma_j}\right)^2$, $m_{l,j} = \sum_{s\in S(l)\setminus\{j\}} m_s$, *and* $\sigma_{l,j}^2 = \sum_{s\in S(l)\setminus\{j\}}\sigma_s^2$. *Then* $\widehat{V}(\mathbf{c})$ *is convex in* $D = [\alpha_1, c_1^M]\times\cdots\times[\alpha_{|L|}, c_{|L|}^M] \subset \Re_+^{|L|}$ *where* $\alpha_l = \max(\max_{j\in S(l)}\alpha_{l,j}, \underline{c}_l)$, *and* $\underline{c}_l = \frac{m_l+\sqrt{m_l^2+4\sigma_l^2}}{2}$.

Proof. Let $\pi_{min} := \min_{s\in S}\pi_s$, then from (10), we have

$$\frac{\partial^2\widehat{V}(\mathbf{c})}{\partial c_l^2} = \frac{d^2\widehat{V}_l(c_l)}{dc_l^2}, = -\left[\frac{\pi_{\min}}{\sigma_{Y_l}^2}\left(\sigma_{Y_l}^2 - c_l(c_l - m_{Y_l})\right)f_{Y_l}(c_l) + \sum_{j\in S(l), j\neq s}(\pi_j - \pi_{\min}) \times \frac{d\left(c_l f_{X_s}(c_l) * g_{S(l)\setminus\{j\}}(c_l)\right)}{dc_l}\right]. \tag{14}$$

Also, $f_{Y_l}(c_l) > 0$ for $\forall c_l \in \Re_+$ and

$$\left(\sigma_{Y_l}^2 - c_l(c_l - m_{Y_l})\right) < 0, \tag{15}$$

if $c_l > \underline{c}_l$. From Lemma 1,

$$\frac{d\left(c_l f_{X_s}(c_l) * g_{S(l)\setminus\{j\}}(c_l)\right)}{dc_l} < 0 \tag{16}$$

if $c_l > \alpha_{l,j}$, where $\alpha_{l,j} = \frac{1}{2}\left[(1-\gamma^2)m_j + 2m_{l,j} + \sqrt{\left((1+\gamma^2)m_j\right)^2 + 4\left(\sigma_j^2+\sigma_{l,j}^2\right)}\right]$, $\gamma_{l,j} = \left(\frac{\sigma_{l,j}}{\sigma_j}\right)^2$, $m_{l,j} = \sum_{s\in S(l)\setminus\{j\}} m_s$, and $\sigma_{l,j}^2 = \sum_{s\in S(l)\setminus\{j\}}\sigma_s^2$. By considering (14)–(16) altogether,

$$\frac{\partial^2\widehat{V}(\mathbf{c})}{\partial c_l^2} \geqslant 0, \tag{17}$$

in $[\alpha_l,\infty]$, where $\alpha_l = \max(\max_{j\in S(l)}\alpha_{l,j}, \underline{c}_l)$. Also,

$$\frac{\partial^2\widehat{V}(\mathbf{c})}{\partial c_l \partial c_m} = 0, \tag{18}$$

when $l \neq m$. Hence, from (17) and (18), $\nabla^2\widehat{V}(\mathbf{c})$ is positive semidefinite for all $\mathbf{c}\in D$ where $D = [\alpha_1, c_1^M]\times\cdots\times[\alpha_{|L|}, c_{|L|}^M] \subset \Re_+^{|L|}$. Consequently, $\widehat{V}(\mathbf{c})$ is convex in D [18]. □

Remark 2. We further investigate the relation between A and D. If we assume that $\sigma_{X_s}/m_{X_s} \ll 1$, which is reasonable for common traffic demand, then we have $\underline{c}_l \approx m_{Y_l}$ and $\alpha_{l,j} \approx m_{Y_l}$, and consequently $D \approx A$. The optimization problem (9) is virtually convex in A with reasonable traffic demand that satisfies $\sigma_{X_s}/m_{X_s} \ll 1$.

Remark 3. From Proposition 3, it is guaranteed that a gradient algorithm based on (10) always converges to a global optimum of (9) if the initial point is in D. Hence, we can obtain the approximate optimal solution $\hat{\mathbf{c}}_l^* = (\hat{c}_l^*, l\in L)$ by solving (10) for each link separately. Even though a necessary optimality condition was derived as a set of coupled equations in [8,9], it is non-trivial to solve these coupled equations, and an iterative algorithm such as a gradient method must be used to find the optimal solution. In applying an iterative algorithm to the optimization problem, the most critical issue is its convergence. However, there was not enough information available on the problem structure such as convexity to guarantee the convergence of an iterative algorithm in [8,9].

Now, we derive several important characteristics that show the effectiveness of the approximation (7).

Proposition 4. *Let* $R(\mathbf{c}) := \widehat{V}(\mathbf{c}) - V(\mathbf{c})$, *then we have*

(i) $R(\mathbf{c}) \geqslant 0$, $\forall\mathbf{c}\in\Re_+^{|L|}$.
(ii) $\lim_{\mathbf{c}\to\infty} R(\mathbf{c}) = 0$.
(iii) $\frac{\partial R(\mathbf{c})}{\partial c_l} \leqslant 0$, $\forall l \in L$.
(iv) $-K_1\max(O_1(c_l), O_2(c_l)) \leqslant \frac{\partial R(\mathbf{c})}{\partial c_l} \leqslant -K_2\min(O_1(c_l), O_2(c_l))$, $\forall l \in L$, *for some positive* K_1 *and* K_2 *that are independent of* c_l. *Here,* $O_1(c_l) = c_l e^{-\frac{(c_l - m_{Y_l})^2}{2\sigma_{Y_l}^2}}$, $O_2(c_l) = e^{-\frac{(c_l - m_{Y_l})^2}{2\sigma_{Y_l}^2}}$.

Proof.

(i) We have

$$R(\mathbf{c}) = \widehat{V}(\mathbf{c}) - V(\mathbf{c}), = \mathbb{E}\left[\sum_s \pi_s X_s\left(\hat{I}_s - I_s\right)\right].$$

Hence, it is sufficient to show that $\hat{I}_s - I_s \geqslant 0$ always holds. We use the following Weierstrass product inequality:

$$1 - \prod_{i=1}^{n}(1-a_i) \leqslant \sum_{i=1}^{n} a_i,$$

where $0 \leqslant a_i \leqslant 1$, $i = 1,\cdots,n$. Since $I_s = \bigvee_{l\in L(s)} I_l(Y_l) = 1 - \prod_{l\in L(s)}(1 - I_l(Y_l))$ and $\hat{I}_s = \sum_{l\in L(s)} I_l(Y_l)$, by simply setting $a_i = I_s$ in the Weierstrass product inequality, we have $I_s = 1 - \prod_{l\in L(s)}(1 - I_l(Y_l)) \leqslant \sum_{l\in L(s)} I_l(Y_l) = \widehat{I}_s$.

(ii) Similarly as in *(i)*, it is sufficient to show that $\lim_{c_l\to\infty,\forall l\in L}(\widehat{I}_s - I_s) = 0$ with probability *1*. When $c_l \to \infty$, $\forall l \in L$, we have $I_l(Y_l) \to 0$, $\forall l \in L$ with probability *1*. Hence, we have $\lim_{c_l\to\infty,\forall l\in L} I_s = \lim_{c_l\to\infty,\forall l\in L}\widehat{I}_s = 0$ with probability *1*. Consequently, $I_l(Y_l) \to 0$, $\forall l \in L$ with probability *1*.

(iii) For any given $k \in L(s)$,

$$I_s = 1 - \prod_{l\in L(s), l\neq k}(1 - I_l(Y_l)) + I_k(Y_k) \times \prod_{l\in L(s), l\neq k}(1 - I_l(Y_l)),$$

and

$$\widehat{I}_s = \sum_{l\in L(s), l\neq k} I_l(Y_l) + I_k(Y_k).$$

Hence,

$$\widehat{I}_s - I_s = I_k(Y_k)\left(1 - \prod_{l\in L(s), l\neq k}(1 - I_l(Y_l))\right) + \sum_{l\in L(s), l\neq k} I_l(Y_l) - 1 + \prod_{l\in L(s), l\neq k}(1 - I_l(Y_l)).$$

Since $\sum_{l\in L(s), l\neq k} I_l(Y_l) - 1 + \prod_{l\in L(s), l\neq k}(1 - I_l(Y_l))$ has no dependency on c_k, we have

$$R(\mathbf{c}) = \int_\Omega \sum_s \pi_s x_s + G(c_1,\cdots,c_{k-1},c_{k+1},\cdots,c_{|L|}),$$

where $\Omega = \bigcap_{l\in L} H_l$ and $H_l = \{\mathbf{x} = (x_s, s\in S) | \sum_{s\in S(l)} x_s > c_l\}$. Since Ω is strictly decreasing for c_k and G has no dependency on c_k, we have $\frac{\partial R(\mathbf{c})}{\partial c_k} \leqslant 0$.

(iv) From (10) and Lemma 1 together with the result in *(iii)*, we have

$$\frac{\partial R(\mathbf{c})}{\partial c_l} = -\left[\lambda_{1,l} c_l f_{Y_l}(c_l) + \lambda_{2,l} f_{Y_l}(c_l)\right], \quad \forall l \in L,$$

for some positive $\lambda_{1,l}$ and $\lambda_{2,l}$ that are independent of c_l. After some manipulation, we can show

$$-\lambda_m \max(O_1(c_l), O_2(c_l)) \leqslant \frac{\partial R(\mathbf{c})}{\partial c_l} \leqslant -\lambda_M \min(O_1(c_l), O_2(c_l)),$$

where

$$O_1(c_l) = c_l e^{-\frac{(c_l - m_{Y_l})^2}{2\sigma_{Y_l}^2}}, \; O_2(c_l) = e^{-\frac{(c_l - m_{Y_l})^2}{2\sigma_{Y_l}^2}},$$

$$\lambda_m = \frac{2}{\sqrt{2\pi}\sigma_{Y_l}} \min(\lambda_{1,l}, \lambda_{2,l}), \text{ and}$$

$$\lambda_M = \frac{2}{\sqrt{2\pi}\sigma_{Y_l}} \max(\lambda_{1,l}, \lambda_{2,l}). \qquad \square$$

Proposition 5. *Let* $\mathbf{c}^*$ *and* $\hat{\mathbf{c}}^*$ *denote solutions to* $\nabla V(\mathbf{c}) = 0$ *and* $\nabla \widehat{V}(\mathbf{c}) = 0$, *respectively. Then, we have* $\hat{\mathbf{c}}^* \succeq \mathbf{c}^*$.

Proof. First, for any c_l, $\forall l \in L$, $\widehat{V}_l(c_l)$ is strictly decreasing in $[\alpha_l, \hat{c}_l^*]$ and strictly increasing in $[\hat{c}_l^*, c_l^M]$ from Proposition 3. Hence, $\frac{\partial \widehat{V}(\mathbf{c})}{\partial c_l} \leqslant 0$ in $[c_l^m, \hat{c}_l^*]$ and $\frac{\partial \widehat{V}(\mathbf{c})}{\partial c_l} \geqslant 0$ in $[\hat{c}_l^*, \infty]$. Since $\frac{\partial V(\mathbf{c})}{\partial c_l} = \frac{\partial \widehat{V}(\mathbf{c})}{\partial c_l} - \frac{\partial R(\mathbf{c})}{\partial c_l}$, $\frac{\partial V(\mathbf{c})}{\partial c_l}|_{c_l = c_l^*} = \frac{\partial \widehat{V}(\mathbf{c})}{\partial c_l}|_{c_l = c_l^*} - \frac{\partial R(\mathbf{c})}{\partial c_l}|_{c_l = c_l^*} = 0$. Hence, $\frac{\partial \widehat{V}(\mathbf{c})}{\partial c_l}|_{c_l = c_l^*} = \frac{\partial R(\mathbf{c})}{\partial c_l}|_{c_l = c_l^*} \leqslant 0$ from Proposition 4 *(ii)*. Consequently, $c_l^* \in [c_l^m, \hat{c}_l^*]$, and we have $c_l^* \leqslant \hat{c}_l^*$. $\square$

From Proposition 4, we can know that $R(\mathbf{c})$ decreases very fast when $\mathbf{c}$ is near $\mathbf{c}^m := (c_l^m, l \in L)$ and changes very slowly as $c_l \to \infty, \forall l \in L$. Consequently, from Proposition 4 together with Proposition 5, we can expect that the approximate optimal solution will give slightly larger bandwidths than the optimal values under a reasonable parameter region. This is a desirable property in a conservative viewpoint.

Also, $\hat{\mathbf{c}}^*$ can be used for further refinement of the solution, i.e., we can find a better solution by restricting our attention to the domain of $D\prime = [\hat{c}_1^* - \epsilon_1, \hat{c}_1^*] \times \cdots \times [\hat{c}_{|L|}^* - \epsilon_{|L|}, \hat{c}_{|L|}^*]$ where $\epsilon_l, \forall l \in L$, are small positive numbers.

As a special case of the problem, if we assume that all the sources are equally treated, i.e., $\pi_1 = \pi_2 = \cdots = \pi_{|S|} := \pi$, then (10) becomes

$$\frac{d\widehat{V}_l(c_l)}{dc_l} = \phi_l - \pi c_l f_{Y_l}(c_l) = 0. \tag{19}$$

From Proposition 3, we have the following corollary:

Corollary 1. *When* $\pi_s = \pi, \forall s \in S$, *there exists a unique vector* $\hat{\mathbf{c}}^* \in A$ *such that* $\widehat{V}(\hat{\mathbf{c}}^*) = \min_{\mathbf{c} \in A} \widehat{V}(\mathbf{c})$ *in* $B = [\underline{c}_1, c_1^M] \times \cdots \times [\underline{c}_{|L|}, c_{|L|}^M]$ *where* $\underline{c}_l = \frac{m_l + \sqrt{(m_l^2 + 4\sigma_l^2)}}{2}$.

Proof. When $\pi_s = \pi$, $\forall s \in S$, from (9), $\widehat{V}(\mathbf{c}) = \sum_l \widehat{V}_l(c_l)$ where $\widehat{V}_l(c_l) = \Phi(c_l) + \int_{c_l}^{\infty} y_l f_{Y_l}(y_l) dy_l$. From Section 2.3, $\widehat{V}_l(c_l)$ is convex when $c_l \geqslant \underline{c} = \frac{m + \sqrt{(m^2 + 4\sigma^2)}}{2}$. Since the sum of a convex function is convex, $\widehat{V}(\mathbf{c})$ is convex. Hence, $\min_{\mathbf{c}} \widehat{V}(\mathbf{c})$ has a unique solution $\hat{\mathbf{c}}^*$ in B. $\square$

Corollary 1 shows the convexity of $\widehat{V}(\mathbf{c})$ in B when $\pi_s = \pi, \forall s \in S$. In fact, from the analysis of the single-link topology in Section 2.3, $\min_{c_l} \widehat{V}_l(c_l)$ has a unique solution $\hat{c}_l \in [c_l^m, c_l^M]$. Hence, $\min_{\mathbf{c}} \widehat{V}(\mathbf{c})$ has a unique solution $\hat{\mathbf{c}}^*$ in A when $\pi_s = \pi, \forall s \in S$.

The overall provisioning algorithm can be summarized as a pseudo code in Algorithm 2, in which $\mathbf{x}_s^{(k)}$ is the realization of X_s in the k-th macro interval, i.e., $\mathbf{x}_s^{(k)} = (x_s^{((k-1)M+1)}, \cdots, x_s^{(kM)})$, $\mathbf{x}^{(k)}$ is the collection of all the $\mathbf{x}_s^{(k)}$'s, $s \in S$, and $\hat{\mathbf{f}}_{\mathbf{X}} = (\hat{f}_{X_s}(x_s), s \in S)$. In the algorithm, it could be a potential issue how to construct the pdf estimate $\hat{\mathbf{f}}_{\mathbf{X}}$ based on the previous data. Here, we simply assume the Gaussian pdf and use the estimate of the mean $\hat{m}_{Y_l}$ and the variance $\hat{\sigma}_{Y_l}$ based on $\mathbf{y}_l^{(k-1)}$ to construct $\hat{f}_{Y_l}(y_l)$ for the k-th macro-time interval. For a non-Gaussian case, we can apply an general estimation approach such as the Parzen window method [19] for the pdf $f_{Y_l}(y_l)$ more accurately.

3. Numerical results

In this section, First, we verify the performance of the optimal solution in a single-link topology. Then, we give numerical results on various topologies to demonstrate the effectiveness of the approximate solution. Finally, we give simulation results that validate the performance of the robust estimator.

Algorithm 2: Bandwidth provisioning algorithm for each line

1:	**for** Each k-th macro interval **do**
2:	Construct $\hat{\mathbf{f}}_{\mathbf{X}}$ based on $\mathbf{x}^{(k)}$
3:	Solve (10) for $\hat{c}_l^*$
4:	$c_l \leftarrow \hat{c}_l^*$
5:	**end for**

3.1. Single-link topology

Here, we deal with the single-link topology to show the performance of the proposed algorithm.

3.1.1. Single-link topology: effects of parameters on $\mathbb{E}[W]$

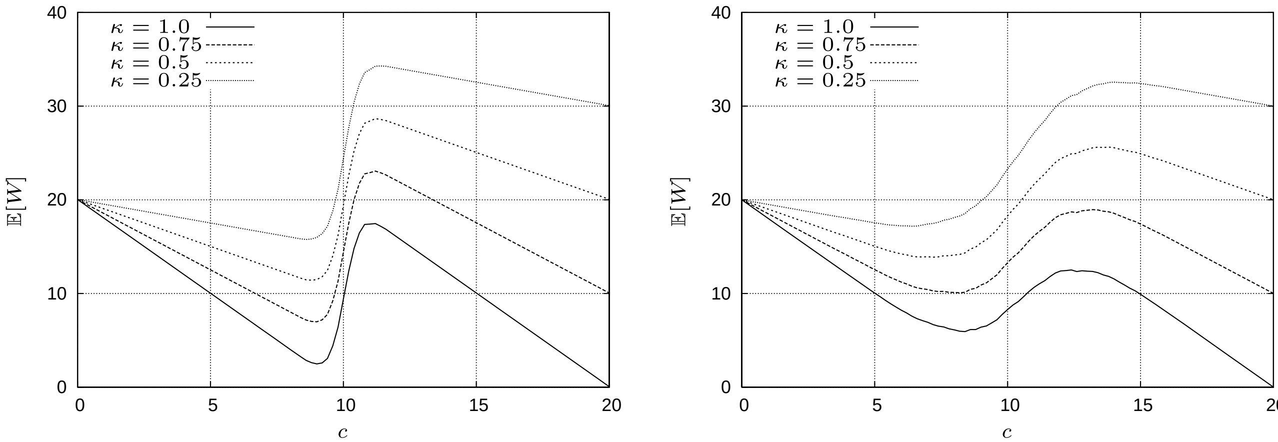
We have conducted simulations to show the effects of parameters on the expected net revenue $\mathbb{E}[W]$ and to compare the performance of the algorithm in Fig. 1 with that of the adaptive online bandwidth provisioning algorithm (ABP) in [8,9]. Here, we have adopted a single-link topology with a single traffic source. We have studied the effects of κ and δ on the expected net revenue $\mathbb{E}[W]$. Fig. 2 show $\mathbb{E}[W]$ as a function of the link bandwidth c for various values of κ and δ. Here, $\mathbb{E}[W]$ are average values over 10 macro-time intervals and $M = 100$. The traffic demand X was generated by the Gaussian distribution with ($m = 10, \sigma = m\delta$). From the figures, we know that the benefit of the appropriate bandwidth provisioning increases as κ decreases. For example, in Fig. 2, the maximum value of $\mathbb{E}[W]$ is smaller than *30* when $\kappa = 0.5$, but larger than *30* when $\kappa = 0.25$. When κ is fixed, we know from Fig. 2 that c^* becomes larger as δ increases. Hence, as long as $\kappa \leqslant \overline{\kappa}$ is satisfied, we need to purchase more bandwidth to maximize the net revenue as δ increases.

3.1.2. Single-link topology: performance analysis under the Gaussian demand

We have compared the performance of the proposed algorithm in Fig. 2 with the ABP algorithm in a single-link topology. Unless otherwise stated, we have set the values of parameters as follows: $g = 4$, $\phi = 1$, and $\pi = 2$ for SON. In the ABP algorithm, we have set $\Theta = \lceil \frac{\overline{X}}{10} \rceil$, $\imath_f = \imath_b = 0.3\Theta$, and $\eta_l = 1$. Note that we have not taken the link utilization threshold η_l into account in the problem formulation, i.e., we have implicitly assumed that the target utilization threshold $\eta_l = 1$ without loss of generality.

We compare the average net revenues of the proposed algorithm and the ABP algorithm. First, we derive the theoretical maximum value of the expected net revenue $W_{max} := \max_c \mathbb{E}[W]$, which can be used as an upper bound of W. The penalty term in W_{max} can be expressed with $c^* = \arg\max_c \mathbb{E}[W]$ as follows:

$$\int_{c^*}^{\infty} x f_X(x) dx = m + \sigma^2 f_X(c^*) - m I_{c^*},$$



(a) $\delta = 0.05$ (b) $\delta = 0.2$

Fig. 2. Average revenue $\mathbb{E}[W]$ vs. link capacity c for various values of the pricing parameter κ.

where $I_{c^*} = \int_0^{c^*} f_X(x)\mathrm{d}x$. Hence,

$$V(c^*) = \phi c^* + \pi\left[\sigma^2 f_X(c^*) + mQ\left(\frac{c^* - m}{\sigma}\right)\right],$$

where $Q(x)$ is the *Q-function* of the Gaussian distribution. Thus, W_{max} is

$$W_{\max} = m\left[g - \pi Q\left(\frac{c^* - m}{\sigma}\right)\right] - \phi c^* - \pi\sigma^2 f_X(c^*). \quad (20)$$

From (20) together with (5), we can obtain the upper bound on the expected net revenue $\mathbb{E}[W]$.

Fig. 3 shows the average net revenues of the proposed algorithm and the ABP algorithm. The maximum bound obtained from (20) is included for comparison. In this simulation, the traffic demand X follows the Gaussian distribution with ($m = 100$, $\sigma = m\delta$). Fig. 3 shows that the net revenue of the proposed algorithm is larger than that of ABP for the entire range of δ, which confirms that the proposed algorithm outperforms ABP in terms of the net revenue. Note that the difference between the net revenue of the proposed algorithm and the maximum bound gets larger as δ increases. The reason is as follows: The proposed algorithm estimates the distribution of the traffic demand based on the data in the previous macro-time interval. However, the size of data used for each pdf estimation is fixed to the number of micro-time slots in one macro-time slot, i.e., $M = 100$ in our simulations. Consequently, the errors caused by pdf estimation will become larger and the performance of the proposed algorithm will degrade as δ increases.

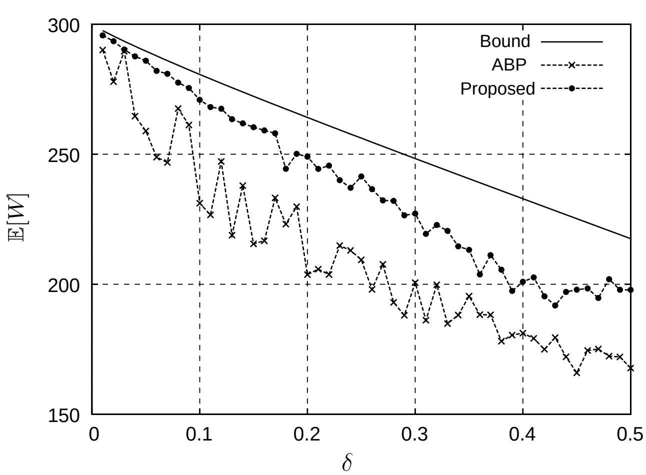


Fig. 3. Average net revenues for ABP, the proposed algorithm, and the theoretical maximum bound with respect to δ, the ratio of the standard deviation to the mean of traffic demand, i.e., $\delta = \sigma/m$.

3.1.3. Single-link topology: performance analysis under various cases of traffic demand

We have performed numerical studies for various cases of different traffic demand. The traffic demand follows the Gaussian distributions with ($m = 100, \sigma = 10 + 5(i-1)$) and ($m = 100 + 20(i-1)$, $\sigma = 10 + 2(i-1)$), for the i-th macro-time slot, in Cases 1 and 2 of Table 1, respectively. So far, we have only considered the cases of the Gaussian traffic demand. Thus, we have performed simulations with the heavy-tailed distribution in Cases 3 and 4 of Table 1. In Case 3, the traffic demand distribution is a single Pareto distribution with the shape parameter $a = \frac{5}{3}$ and the scale parameter $b = 40$. Further, in Case 4, we have taken into account an aggregation of ten Pareto sources. The mean of each source is set to *10*, and the shape and the scale parameters for the i-th source are set to $a = \frac{10}{10 - 2\lceil\frac{i}{2}\rceil}$ and $b = 2\lceil\frac{i}{2}\rceil$, respectively. Table 1 shows that the net revenue of the proposed algorithm is larger than that of ABP for all the cases. It also shows the *relative gain*, which is the ratio of the net revenues between the proposed algorithm and ABP. Especially, in Case 3, the proposed algorithm shows outstanding performance improvement compared to ABP. Hence, we can expect that the proposed algorithm works well in bursty traffic demand.

3.2. More general topologies

Now, we consider more general topologies such as a two-link topology, a parking-lot scenario, and a tree topology.

3.2.1. Illustrative example: a simple two-link topology

We need to verify the accuracy of the approximate optimal solution obtained from (10). For this purpose, we introduce a simple, yet illustrative two-link network. Consider a network with two successive links, l_1 and l_2 as in Fig. 4. Let X_1 and X_2 denote two

Table 1
Average net revenue $\mathbb{E}[W]$ for various traffic demand distributions

	Case 1	Case 2	Case 3	Case 4
ABP	186	349	82	138
Proposed	214	435	116	157
Relative gain	115%	125%	141%	114%

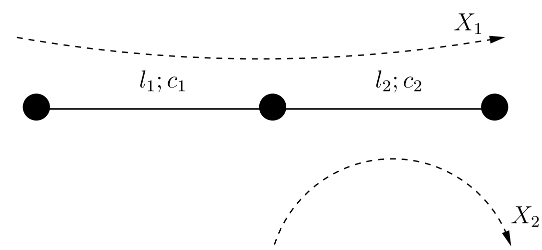


Fig. 4. A two-link topology.

traffic flows, each of which traverses $L(s_1) = \{l_1, l_2\}$ and $L(s_2) = \{l_2\}$, respectively. The net revenue W is

$$W = g_1X_1 + g_2X_2 - (\phi_1c_1 + \phi_2c_2) - (\pi_1Z_1 + \pi_2Z_2).$$

Here, $Z_1 = X_1I_1(X_1,X_2)$ and $Z_2 = X_2\,I_2(X_1,X_2)$ where $I_1(X_1,X_2)$ and $I_2(X_1, X_2)$ are indicator functions as follows:

$$I_1(X_1,X_2) = \begin{cases} 1 & \text{if } X_1 \geqslant \min(c_1, c_2 - X_2), \\ 0 & \text{otherwise}, \end{cases}$$

and

$$I_2(X_1,X_2) = \begin{cases} 1 & \text{if } X_2 \geqslant c_2 - X_1, \\ 0 & \text{otherwise}, \end{cases}$$

respectively. The objective function becomes

$$\text{minimize } V(c_1,c_2) = \phi_1c_1 + \phi_2c_2 + \mathbb{E}[\pi_1Z_1 + \pi_2Z_2]$$
$$\text{subject to } c_l \geqslant m_{Y_l}, l = 1,2.$$

By introducing (7), we have the following approximate objective function $\widehat{V}$:

$$\widehat{V}(c_1,c_2) = \phi_1c_1 + \phi_2c_2 + \pi_1\int_{c_1}^{\infty} x_1f_{X_1}\mathrm{d}x_1 + \int_0^{\infty}\int_{c_2-x_2}^{\infty}(\pi_1x_1 + \pi_2x_2)f_{X_1}\mathrm{d}x_1f_{X_2}\mathrm{d}x_2.$$

The necessary optimality conditions are

$$\frac{\partial\widehat{V}}{\partial c_1} = \phi_1 - \pi_1c_1f_{Y_1}(c_1) = 0,$$
$$\frac{\partial\widehat{V}}{\partial c_2} = \phi_2 - \pi_2c_2f_{Y_2}(c_2) - (\pi_2 - \pi_1)\left[c_2f_{X_1}(c_2) * f_{X_2}(c_2)\right] = 0,$$

where $Y_1 = X_1$ and $Y_2 = X_1 + X_2$. Hence, $f_{Y_1} = f_{X_1}$ and $f_{Y_2} = f_{X_1} * f_{X_2}$.

Now, we show numerical results with the following traffic demand: $X_s \sim \mathcal{N}(m_s, \sigma_s^2)$, $s = 1,2$ where $m_1 = 10^3$, $m_2 = 2 \cdot 10^3$ and $\sigma_1 = 100$, $\sigma_2 = 250$. The values of parameters used in the simulation are $\phi_1 = \phi_2 = 1$ and $\pi_1 = 3.0$, $\pi_2 = 1.5$. The optimal bandwidth and the corresponding V are $\mathbf{c}^* = (c_1^*, c_2^*) = (1177, 3604)$ and $V(\mathbf{c}^*) = 4.853 \cdot 10^3$ while the approximate solution and the corresponding V are $\hat{\mathbf{c}}^* = (1177, 3605)$ and $V(\hat{\mathbf{c}}^*) = 4.854 \cdot 10^3$. Hence, we know that the proposed approximate problem gives a very accurate solution to the original problem. From Fig. 5, we can also verify that the approximation error decreases very fast the link bandwidths increases. We specify $\mathbf{c}^*$ and $\hat{\mathbf{c}}^*$ in Fig. 5 (b) as '*o*' and '*x*', respectively. As shown in Fig. 5 (b), $V - \widehat{V}$ is negligible near $\mathbf{c}^*$ and $\hat{\mathbf{c}}^*$.

3.2.2. Parking lot scenario

Now, for further investigation on the performance of the proposed algorithm, consider the so called "parking lot" scenario in Fig. 6. There is a traffic source traversing all the links in an end-to-end manner. Each link also has one cross traffic which uses that link only. We obtain the approximate optimal solution by solving (10) with a gradient algorithm. For comparison, we also derive the optimal solution by numerical enumeration. The values of parameters used for simulations are as follows: $\phi_l = 1$ for all $l \in L$,

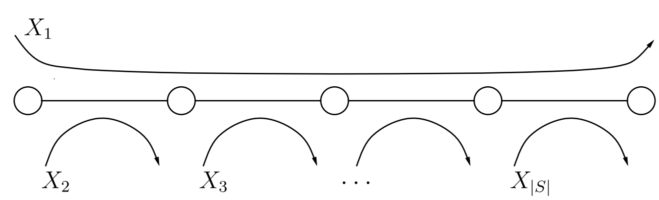
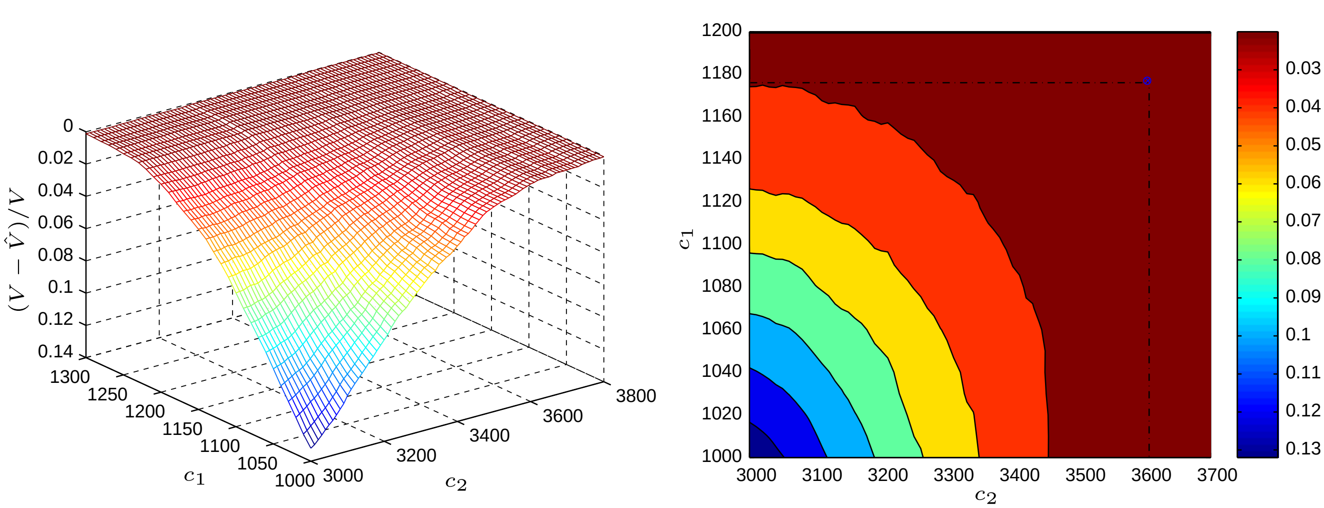


Fig. 6. A parking lot scenario.

Table 2
Performance of the proposed algorithm under parking lot scenarios

$\lvert L\rvert$	$\mathbf{c}^*$	$\hat{\mathbf{c}}^*$	Relative error in % ($= (\hat{\mathbf{c}}^* - \mathbf{c}^*)/\mathbf{c}^*$)
3	(23.1,23.5,23.2)	(23.5,23.7,23.5)	(1.73,0.85, 1.29)
4	(23.6,23.1,23.5,23.5)	(23.7,23.5,23.6,23.6)	(0.42,1.73,0.43,0.43)
5	(23.5,23.2,23.5,23.5,23.3)	(23.7,23.4,23.6,23.7, 23.4)	(0.85,0.86,0.43,0.85,0.43)



(a) Relative error $(V - \hat{V})/V$

(b) Contour plot of the relative error

Fig. 5. Relative error $(V - \widehat{V})/V$ vs. link capacities c_1 and c_2 under the two-link topology in Fig. 4.

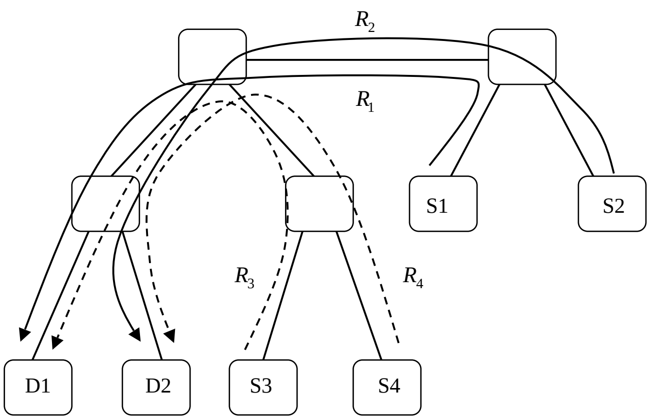


Fig. 7. A tree topology.

$\pi_s = 3$, $m_s = 10$, $\sigma_s = 1$ for all $s \in S$, and $\gamma = 0.5$, $\epsilon = 0.001$. From Table 2, we can see the effectiveness of the proposed algorithm. The error of our solution is less than 2% in all the cases.

3.2.3. Tree topology

To consider a more practical situation, we adopt the same tree topology in [8], which is shown in Fig. 7. There are four traffic sources s_i, $i = 1,2,3,4$, each of which traverses route R_i, $i = 1,2,3,4$, as given in Fig. 7. For each source i, X_i follows the Gaussian traffic distribution with the mean of $m = 100$, and σ is set to *10* for s_1 and s_2, and *20* for s_3 and s_4. In this simulation, we set $g_s = 10$ and $\pi_s = 2$ for all s as in [8,9]. The simulation is performed for 10 macro-time intervals. The average net revenue of the proposed algorithm is *2107* and that of ABP is *1906*. Hence, the proposed algorithm gives 14% more net revenue than ABP.

4. Conclusion

In this paper, we studied the problem of revenue optimization via bandwidth provisioning in SON. We rigorously formulated the problem in an optimization framework taking into consideration the stochastic processes involved. First, we derived a necessary and sufficient optimality condition in a single-link topology and suggested a gradient algorithm for obtaining the optimal solution. Then, we showed that the problem in a general topology can be nicely approximated by a separable convex optimization problem. The convergence of a gradient algorithm was guaranteed by the convexity of the approximate problem. Also, the curse of dimensionality problem was circumvented by the separability of the problem. We further derived several important characteristics of the objective function, which helps to show the effectiveness of the approximate optimal solution. In a reasonable parameter region, the approximate optimal solution is shown to be slightly larger than the optimal solution. We further extended the problem and derived a first-order necessary optimality condition for non-stationary traffic demand. Finally, to incorporate the dynamic bandwidth provisioning, we introduced a robust estimator for non-stationary traffic demand.

Acknowledgement

This work has been supported in part by the BK21 IT Program of the Korean Ministry of Education, Science and Technology.

References

[1] J. Jannotti et al., Overcast: reliable multicasting with an overlay network, in: Proceedings of the 4th Symposium on Operating Systems Design and Implementation (OSDI), 2000.
[2] Y. Chawathe, Scattercast: an architecture for internet broadcast distribution as an infrastructure service, Ph.D. Thesis, Computer Science Department, University of California, Berkeley (December 2000).
[3] S. Savage et al., Detour: a case for informed internet routing and transport, IEEE Micro 19 (1) (1999) 50–59.
[4] D.G. Andersen, H. Balakrishnan, M. Kaashoek, R. Morris, Resilient overlay networks, in: Proceedings of the 18th ACM Symposium on Operating Systems Principles (SOSP) 2001.
[5] M. Waldvogel, R. Rinaldi, Efficient topology-aware overlay network, ACM Computer Communication Review 33 (1) (2003) 101–106.
[6] Virtela Communications. Available from: <http://www.virtela.com>.
[7] Internap Network Services Corporation. Available from: <http://www.internap.com>.
[8] Z. Duan, Z.-L. Zhang, Y.T. Hou, Service overlay networks: SLAs, QoS and bandwidth provisioning, in: Proceedings of IEEE International Conference on Network Protocol (ICNP), 2002.
[9] Z. Duan, Z.-L. Zhang, Y.T. Hou, Service overlay networks: SLAs, QoS, and bandwidth provisioning, IEEE/ACM Transactions on Networking 11 (6) (2003) 870–883.
[10] D. Mitra, Q. Wang, Stochastic traffic engineering, with applications to network revenue management, in: Proceedings of IEEE INFOCOM, 2003.
[11] A. Jüttner, I. Szabó, Áron Szentesi, On bandwidth efficiency of the hose resource management model in virtual private networks, in: Proceedings of IEEE INFOCOM, 2003.
[12] B. Krithikaivasan, K. Deka, D. Medhi, Adaptive bandwidth provisioning envelope based on discrete temporal network measurement, in: Proceedings of IEEE INFOCOM, 2004.
[13] H.T. Tran, T. Ziegler, Adaptive bandwidth provisioning with explicit respect to QoS requirements, in: Proceedings of the Workshop on Quality of Future Internet Services (QoFIS), 2003.
[14] C.-K. Tham, T.C.-K. Hui, Reinforcement learning-based dynamic bandwidth provisioning for quality of service in differentiated services networks, Elsevier Computer Communications 28 (15) (2005) 1741–1751.
[15] M. Marchese, M. Mongelli, Adaptive rate allocation and resource planning for service level agreement maintenance in satellite communications, Elsevier Computer Communications 30 (16) (2007) 3179–3192.
[16] H. Berg, M. Mandjes, R. Meent, A. Pras, F. Roijers, P. Venemans, QoS-aware bandwidth provisioning for IP network links, Elsevier Computer Networks 50 (5) (2006) 631–647.
[17] L. Song, U. Deshpande, U.C. Kozat, D. Kotz, R. Jain, Predictability of WLAN mobility and its effects on bandwidth provisioning, in: Proceedings of IEEE INFOCOM, 2006.
[18] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, 1999.
[19] E. Parzen, On estimation of a probability density function and mode, Annals of Mathematical Statistics 33 (1962) 1065–1076.