Pricing for Revenue Maximization in Inter-DataCenter Networks

Zhenzhe Zheng*, R. Srikant†, Guihai Chen
Department of Computer Science and Engineering, Shanghai Jiao Tong University, China
†Coordinate Science Lab, Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana, IL, USA
Email: zhengzhenzhe@sjtu.edu.cn, rsrikant@illinois.edu, gchen@cs.sjtu.edu.cn

***Abstract*—As more applications and businesses move to the cloud, pricing for inter-datacenter links has become an important problem. In this paper, we study revenue maximizing pricing from the perspective of a network provider in inter-datacenter networks. Designing a practical bandwidth pricing scheme requires us to jointly consider the requirements of envy-freeness and arbitrage-freeness, where envy-freeness guarantees the fairness of resource allocation and arbitrage-freeness induces users to truthfully reveal their data transfer requests. Considering the non-convexity of the revenue maximization problem and the lack of information about the users' utilities, we propose a framework for computationally efficient pricing to approximately maximize revenue in a range of environments. We first study the case of a single link accessed by many users, and design a $(1+\epsilon)$-approximation pricing scheme with polynomial time complexity and information complexity. Based on dynamic programming, we then extend the pricing scheme for the tollbooth network, preserving the $(1+\epsilon)$ approximation ratio and the computational complexity. For the general network setting, we analyze the revenue generated by uniform pricing, which determines a single per unit price for all potential users. We show that when users have similar utilities, uniform pricing can achieve a good approximation ratio, which is independent of network topology and data transfer requests. The pricing framework can be extended to multiple time slots, enabling time-dependent pricing.**

I. Introduction

With the growing deployment of globally-distributed applications, inter-datacenter networks are becoming an important resource in cloud computing. Many services rely on low-latency communication and high-throughput data transfers among data centers to improve the quality of experience for their consumers. For example, media companies can deliver high definition video content to consumers in multiple areas [1], search engines can synchronize search indexes between data centers [2], and big data applications can move datasets collected at one data center to another for further analysis. Considering the emerging market demand for inter-datacenter bandwidth, online service providers, such as Amazon, Google, and Microsoft, have built dedicated wide area networks (WANs) to connect their datacenters [2], [3], and launched data transfer services like AWS direct connect [4], to facilitate the success of globally-distributed applications.

*This work was done while the author visited Coordinate Science Lab, University of Illinois at Urbana-Champaign, Urbana, IL, USA

This work was supported in part by the State Key Development Program for Basic Research of China (973 project 2014CB340303), by NSF Grants CIF-1409106 and CPS ECCS-1739189, in part by China NSF grant 61672353 and 61472252.

In this paper, we study the pricing for network connectivity in inter-datacenter networks from a network provider's point of view. To access inter-datacenter networks, users submit data transfer requests, containing source and destination data centers and transmission time intervals, to the network provider. The network provider designs a dynamic pricing scheme to specify the bandwidth charges to potential users. Users response to the charges by choosing a certain data rate to transmit data over the network. A recent work [5] has considered the objective of maximizing social welfare, which measures the aggregated "happiness" of network provider and users. This objective may not be in the interest of network provider, especially in public WANs, where the network provider wants to extract more revenue to maximize the return on investment in inter-datacenter WAN. Thus, our focus here is on designing pricing scheme for revenue maximization.

A practical revenue maximizing pricing scheme should satisfy the properties of *envy-freeness* and *arbitrage-freeness*. In economics, envy-freeness is a natural fairness criterion, requiring that given a pricing rule, users would be allocated the resources, *i.e.*, data transmission rates in our context, that maximize their utilities [6]. We further extend this criterion to avoid other possible unfairness among users. We require that users consuming a set of resources, should be charged no less than the users consuming only a subset of the resources, as they use more resources. In the context of inter-datacenter networks, this requires the per unit price for users to transmit data over a certain route, should be no smaller than that to users, who only use a sub-path of the route. In designing pricing schemes, the network provider also needs to consider the potential arbitrage behaviors from users. One simple arbitrage behavior may be purchasing a sequence of single resources rather than buying a bundle of resources at one time, if the total price of individual resource is less than the price of the bundle. Suppose a user requires data transfer over a route $A \rightarrow B \rightarrow C$. An intelligent user may have an incentive to divide her request into two sub requests over routes $A \rightarrow B$ and $B \rightarrow C$, respectively, if the price of the whole request is larger than the sum of prices of the two sub requests. Hence, designing a pricing scheme with the guarantee of envy-freeness and arbitrage-freeness can avoid discontent among users and induce users to truthfully reveal their requests.

To design an efficient revenue maximizing pricing scheme, we have to consider the computational complexity from the perspectives of both time complexity and information complexity. Existing works only examined the special cases of revenue maximization problem, such as adopting specific utility functions for users [7], [8] or considering a simplified network topology [9]–[11], to derive positive results. However, the revenue maximization problem in general is non-convex, and thus it is computationally intractable to calculate the optimal prices. Therefore, we are interested in designing computationally efficient pricing to achieve approximate revenue maximization in general scenarios. Different from the previous works in the complete information setting [8], [12], we attempt to maximize revenue without full knowledge of users' utility functions, taking only a limited amount of communication with users. We measure the information complexity by the number of rounds of interaction needed to achieve the optimal revenue within some given approximation factor. Considering the high computational complexity of deriving optimal revenue, the network provider faces a trade-off between achieving optimal revenue and keeping pricing schemes efficient.

We now summarize the main contributions in this paper.

- We model the interaction between network provider and users: the network provider sets a per unit price for each user and the user responds to the price by choosing a certain data rate to transmit data. We consider two types of envy-free and arbitrage-free pricing: *item pricing* and *uniform pricing*, to determine the prices, and investigate their revenue guarantees within a range of environments.
- We begin with studying a single link case, in which the network provider determines the price of link to maximize revenue. Using a straightforward idea, we can design a $(1+\epsilon)$-approximation pricing scheme with polynomial time complexity and information complexity. This result is the basis of developing solutions for more complex network cases.
- Motivated by the current pricing strategy in inter-datacenter networks, we then consider a tollbooth case with a tree network topology. Based on the idea of dynamic programming, we propose an item pricing scheme to determine the price for each link, and charge a user with the sum of prices of links in her route. Our analysis shows that this item pricing scheme preserves the $(1+\epsilon)$ approximation ratio and the computational complexity.
- For general network setting, we investigate two variants of item pricing, and show that they are impractical due to either high computational complexity or loose revenue guarantee. Our main result shows that uniform pricing, which charges a single per unit price to all users, satisfies the requirements of envy-freeness and arbitrage-freeness, and has good revenue guarantee when users have similar utility functions.
- We further extend pricing schemes to multiple time slots, enabling time-dependent pricing. We show that the pricing schemes preserve the approximation ratio and only introduce an additional constant factor for computational complexity.
- Finally, we evaluate the performance of uniform pricing with specific utility functions in a real inter-datacenter networks. The evaluation results show that uniform pricing outperforms a trivial item pricing, and the ratio between the optimal revenue and the obtained revenue is less than 1.6.

II. Related Work

Pricing for communication networks has been well studied in the literature. The Transmission Control Protocol (TCP) can be viewed as a congestion pricing scheme [13], [14], and new network protocol can also be reverse-engineered as a pricing-based solution to a Network Utility Maximization (NUM) [15] with specific utility functions. The objective in NUM is to maximize the sum of users' utilities, which can be considered as the social welfare of network. Our work differs from this line of works, because our goal is to maximize revenue, which is non-convex in general and is more challenging.

There is an extensive literature investigating revenue maximization with various pricing schemes in communication network [7], [8], [10], [12], [16]. Acemoglu *et al.* studied entry price strategy and power allocation rule in wireless network [16]. Başar and Srikant designed usage-based pricing schemes in a single link network [8] and a multiple-link network [7], and derive the condition under which expanding network capacity increases revenue. Shakkottai *et al.* proposed the concept of price of simplicity to measure the revenue loss due to using flat entry pricing in single link network [9]. Our result can be considered as the extension of this concept to multiple-link network. Our work differs from the above works due to the factor that we consider general utility functions and network topologies in incomplete information setting. Most of the above works either adopted specific utility functions, such as logarithmic function [7], [8], or considered a simplified network topology, such as single link network [9]–[11]. The problem of revenue maximization has also been extensively studied in economics and theoretical computer science [6], [17], [18]. However, most of these works focused on discrete linear utilities [6], [17], rather than the continuous concave utilities considered in this work. The work [18] needs to know the distribution of users' utilities to maximize revenue.

Employing software defined networking (SDN) technique, large cloud companies have built dedicated WANs, such as Google's B4 [2] and Microsoft's SWAN [3], to connect their geo-distributed data centers. Kandula *et al.* presented Tempus, a WAN data transfer framework, to schedule long-running data transfers in spatial and temporal dimension [19]. Kumar *et al.* presented Google's Bandwidth Enforce (BwE) for WAN bandwidth allocation, enabling WAN to run at a high level of utilization [20]. These works enable network provider to improve the utilization of inter-datacenter bandwidths from the perspective of engineering, our work, from the perspective of economics, attempts to design a revenue maximizing pricing scheme for network provider to extract more revenue from inter-datacenter bandwidth market.

III. Preliminaries

A. System Model

A Network Provider: We consider a network provider managing a WAN $G=(V,E)$ of interconnected datacenters. Each node $v \in V$ denotes a data center, and each edge $e=(v,w)$ represents a WAN link connecting datacenters v and w or an ingress/egress link between datacenter v and an ISP w. We discretize time into T time slots, where each time slot corresponds to a fixed time interval, such as five minute interval. The provider associates a price $p_{e,t}$ for each link $e \in E$ at each time slot t, meaning that users have to pay $p_{e,t}$ for transferring data with unit data rate over link e at time slot t. We use matrix $\boldsymbol{P} = (p_{e,t}) \in \mathbb{R}^{|E|\times T}$ to denote the prices for different links at different time slots. We assume that links do not have explicit capacity constraints, and adjust demands over links through the control of prices on the links [21].

Network Users: There is a set of N network users, denoted by $\mathcal{N} = \{1, 2, \cdots, N\}$, who make data transfer requests. The request from user $i \in \mathcal{N}$ can be specified as a tuple $\{s_i, d_i, t_i^1, t_i^2, u_i(\cdot)\}$. The nodes s_i and d_i represent the source and destination of data transfer, respectively, meaning that the data can be transmitted along a set of possible paths (or routes) R_i from s_i to d_i. For convenience of discussion, we only express explicit solutions to the single-minded case, in which each of user $i \in \mathcal{N}$ pre-determines a specific route r_i from R_i, to transmit data.[1] We consider a data rate request scenario, in which each of users requires a constant data transmission rate over a certain period of time. The user $i \in \mathcal{N}$ also specifies a time interval $[t_i^1, t_i^2]$, where t_i^1 corresponds to the starting time slot to transmit data and t_i^2 represents the ending time slot. We assume that the lengths of time intervals are bounded by a constant $\bar{t}$, *i.e.*, $t_i^2 - t_i^1 \leq \bar{t}, \forall i \in \mathcal{N}$. The user $i \in \mathcal{N}$ has a (private) utility $u_i(x_i)$, which measures the worth of the data transfer at data rate x_i over the time interval $[t_i^1, t_i^2]$. Throughout the paper, we adopt the following assumption on the utility function $u(x)$.

Assumption 1. *The utility function* $u : [0, \infty] \to [0, \infty]$ *is an increasing, strictly concave, and differentiable function.*

Charge: The network provider charges a payment $p_i(x_i)$ to user $i \in \mathcal{N}$ for using the network with a data rate x_i during the time interval $[t_i^1, t_i^2]$. In general, the price of a data transfer is related to the allocated data rate, the source and destination nodes, the requested transmission time interval, the market competition for resources, etc. For simplicity, we consider a linear usage-based pricing: the charge to a user is proportional to the amount of allocated data rate, *i.e.*, $p_i(x_i) = p_i \times x_i$. For convenience of discussion, we normalize the feasible prices for p_i into the interval $[1, \bar{p}]$. To determine the per unit price p_i, we further investigate two different pricing schemes under the requirements of envy-freeness and arbitrage-freeness:

• *Item Pricing Scheme* [6], [17]: We can set the per unit price p_i for user $i \in \mathcal{N}$ during the time interval $[t_i^1, t_i^2]$ as $p_i \triangleq \sum_{t_i^1 \leq t \leq t_i^2} p_{i,t}$, where $p_{i,t}$ is the unit price for user i at time slot t, and is equal to the total price of links over the selected route r_i, *i.e.*, $p_{i,t} \triangleq \sum_{e\in r_i} p_{e,t}$, leading to

$$p_i = \sum_{t_i^1 \leq t \leq t_i^2} \sum_{e\in r_i} p_{e,t}. \tag{1}$$

• *Uniform Pricing Scheme*: We also consider the simplest usage-based pricing scheme: uniform pricing, which charges a single per unit price for all potential users, independent of their selected routes and requested time intervals.

[1] Our solution can also handle the multi-minded case, in which users can transmit data over multiple paths simultaneously.

Demand: The net utility of user i can be expressed as the difference between her utility and payment over the allocated data transmission rate, *i.e.*, $u_i(x_i) - p_i \times x_i$. Given the price quote p_i, each user $i \in \mathcal{N}$ determines her demand for data rate by solving a utility maximization problem:

$$\text{Maximize} \quad u_i(x_i) - p_i \times x_i \quad \text{over} \quad x_i \geq 0. \tag{2}$$

Since $u_i(x_i)$ is strictly concave by Assumption 1, the utility maximization problem admits a unique solution x_i^* such that

$$u_i'(x_i^*) = p_i, \quad \text{if } x_i^* > 0, \quad \text{and} \quad u_i'(x_i^*) < p_i \quad \text{if } x_i^* = 0.$$

We can also express the demand of user i with the per unit price p_i as:

$$D_i(p_i) \triangleq \left[u_i'^{-1}(p_i)\right]_+, \tag{3}$$

where $\left[u_i'^{-1}(p_i)\right]_+$ denotes $\max(u_i'^{-1}(p_i), 0)$. Since the utility function $u_i(x)$ is strictly concave, $u_i'(x)$ is a decreasing function, and thus the demand function $D_i(p_i)$ is non-increasing with respect to price p_i over the range $[0, u_i'(0)]$.

B. Problem Formulation

We associate with the network provider a revenue maximization problem to determine the optimal prices to charge, which can be formally formulated as follows:

Problem: *Revenue Maximization with Item Pricing*
Objective: Maximize $\quad L(\boldsymbol{P}) = \sum_{i\in\mathcal{N}} p_i \times x_i$
Subject to:

$$\begin{aligned} p_i &= \sum_{t_i^1 \leq t \leq t_i^2} \sum_{e\in r_i} p_{e,t}, && \forall i \in \mathcal{N}, \\ x_i &= D_i(p_i), && \forall i \in \mathcal{N}, \\ 1 \leq p_i &\leq \bar{p}, && \forall i \in \mathcal{N}. \end{aligned}$$

The objective of the optimization problem is to maximize the overall revenue $L(\boldsymbol{P})$ from N users. The decision variable $p_{e,t} \in \boldsymbol{P}$ denotes the unit price of link e at time slot t. The first set of constraints states the unite price to each user given the price matrix $\boldsymbol{P} = (p_{e,t})$, when we adopt item pricing. The second set of constraints is the requirement of *envy-free* rate allocation [6], meaning that each user receives the data rate equal to her demand. The third set of constraints states the feasible range of unite prices to users.

Solving the above revenue maximization problem is non-trivial because of the following three major challenges. The first challenge comes from the non-convexity of the objective function. The revenue function $p_iD_i(p_i)$ (or $x_iu_i'(x_i)$) could be non-convex in general. The second challenge comes from the incomplete information of the network provider. In large markets, the network provider may not have full knowledge of users' utilities, and can only observe the responses (or demands) of users given certain prices. The last challenge is the complex price correlation in both link and temporal dimensions. The adjustment of the price on one single link at a specific time slot has a/an direct/indirect effect on the charge to potentially all users, due to coupling of users through shared links and common time intervals. Jointly considering the above challenges, we propose a framework of pricing mechanisms to approximate the maximum revenue (in the sense of bounded approximation ratios) under different scenarios. We formally define the revenue loss metric used in this paper.

Definition 1 (**α-Approximation Pricing Scheme**). *A pricing scheme is called α-approximation if the ratio between optimal revenue and the revenue obtained by the pricing scheme is less than or equal to α for all instances of the problem. α is also called the approximation ratio of the pricing scheme.*

We measure the computational complexity of pricing schemes from time complexity and information complexity. The information complexity measures the smallest number of rounds of interaction (between network provider and users) needed to maximize revenue within a desired approximation ratio [22].

IV. Pricing Schemes for a Special Case

In this section, we first consider a special case of the problem formulated above, where there is only one time slot, *i.e.*, $T=1$. The assumption $T=1$ is equivalent to considering a problem where all users transmit for the same duration of time, and thus, is an important special case to study in its own right. In this special case, we only need to determine the prices of links, not necessary to consider price constraints in the temporal dimension. However, it is still non-trivial to derive a complete solution for the general network setting. We begin with designing pricing scheme for the simplest single link case, and then extend it to adapt to the tollbooth case, in which the network topology is restricted to a tree. We finally derive some positive results for the general network topology under some assumptions on demand functions.

A. Single Link Case

In single link case, the network provider only needs to determine a single price for the link to maximize revenue, further reducing the complexity of price correlation. Hence, we have a simple version of revenue maximization:

$$\text{Maximize} \quad L(p) = \sum_{i\in\mathcal{N}} p \times D_i(p) \quad \text{over} \quad 1 \le p \le \bar{p}. \tag{4}$$

In general, this optimization problem remains to be non-convex, and the demand functions are also unknown to the network provider. By leveraging the specific structure of this optimization problem, we propose a simple and efficient pricing scheme to achieve good revenue guarantee. We define a vector of candidate discrete prices as

$$\hat{\boldsymbol{p}} \triangleq (\hat{p}_0, \hat{p}_1, \cdots, \hat{p}_K), \tag{5}$$

where $\hat{p}_k = (1+\epsilon)^k$ for $k = 0, 1, \cdots, K$ and a constant parameter $\epsilon > 0$. Since the upper bound of feasible prices is $\bar{p}$, we can set $K = \left\lfloor \frac{\log(\bar{p})}{\log(1+\epsilon)} \right\rfloor$. The specific pricing scheme contains two steps:

- Step 1: For each candidate price $\hat{p}_k$, we calculate the total revenue $L(\hat{p}_k) = \sum_{i\in\mathcal{N}} \hat{p}_k \times D_i(\hat{p}_k)$.
- Step 2: We select the candidate price $\hat{p}_k$ with the maximum revenue $L(\hat{p}_k)$ as the final price.

Suppose the selected price is $\hat{p}^*$, and the corresponding revenue is $ALG = L(\hat{p}^*)$. We denote the optimal revenue as $OPT = L(p^*)$, where p^* is the optimal price. Using a straightforward analysis, we can characterize the revenue guarantee of the pricing scheme in the following theorem.

TABLE I
Approximation ratio vs. computational complexity ($\bar{p} = 100$)

Approximation ratio $1+\epsilon$	1.1	1.2	1.5	2	3
Number of interactions per user K	48	25	11	6	4

Theorem 1. *For the single link case, we can achieve an approximation ratio of $(1+\epsilon)$, with polynomial time complexity and information complexity $O(NK)$, where N is the number of users and $K = \left\lfloor \frac{\log(\bar{p})}{\log(1+\epsilon)} \right\rfloor$.*

Proof. For the optimal price p^*, there exists some integer $k \in [0, K]$ such that $(1+\epsilon)^k \le p^* \le (1+\epsilon)^{k+1}$. Thus, we can get

$$\begin{aligned} OPT &= p^* \sum_{i\in\mathcal{N}} D_i(p^*) \le (1+\epsilon)^{k+1} \sum_{i\in\mathcal{N}} D_i(p^*) \\ &\le (1+\epsilon) \times (1+\epsilon)^k \sum_{i\in\mathcal{N}} D_i\left((1+\epsilon)^k\right) \qquad (6) \\ &\le (1+\epsilon) \times ALG. \qquad (7) \end{aligned}$$

The inequality (6) comes from the fact that $D_i(p)$ is non-increasing with respect to price p over the range $[1, \bar{p}]$. The inequality (7) holds due to Step 2 in the pricing scheme.

Since we have to calculate the revenue of all candidate prices, each of which takes $O(N)$ time, the time complexity is $O(NK)$. The network provider needs to interact with each user for K times, to learn her demands under different candidate prices. Thus, the information complexity is $O(NK)$. $\square$

It is worth emphasizing that the pricing scheme works for general demand functions, only requiring the demand to be a non-increasing function of the price. Another interesting observation can be made from the theorem above is that the pricing scheme achieves a good approximation ratio through only evaluating the demand functions a polynomial number of times, without full knowledge of demand functions. In other words, the network provider only needs to interact with each user K times to obtain a $(1+\epsilon)$ approximation ratio.[2] We show the trade-off between approximation ratio and computational complexity in Table I when the upper bound of prices is 100.

B. Tollbooth Case

We now extend the pricing scheme in single link case to a more complex network called tollbooth network. In the tollbooth case, the network topology is a tree, and users share one common source, which we consider as the root of the tree. This case is motivated by the current pricing strategy in inter-datacenter networks, where the network provider designs different price quotes for data transfer requests from different data centers. We illustrate the tollbooth case in Figure 1, which is a minimum spanning tree of Google's inter-deatacenter networks in the United States [2].

[2] We assume users are price takers and will truthfully reveal their demands given a price. We leave the consideration of strategic behaviors of users [23] to our future work.

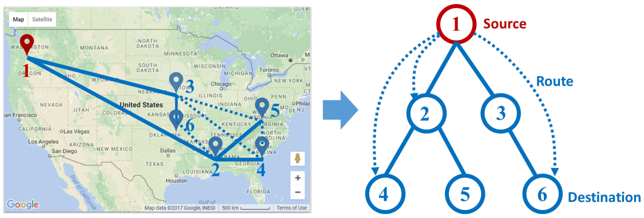


Fig. 1. A tollbooth network, which is a minimum spanning tree of Google's Inter-datacenter networks in the United States [2].

We propose a pricing scheme for the tollbooth case based on dynamic programming. For each node w in the tree, we use $\mathcal{N}_w$ to denote the set of users, who request data transfer over the route r_w from the root to the destination node w. We also denote the children of node w by a set V_w. We define $L(w, p_w)$ to be the maximum revenue that we can obtain from the users, whose destinations are located in the subtree Tr_w, given that the route r_w has an exact price p_w. We are then interested in calculating $L(w_r, 0)$, where node w_r is the root of tree. Using these terminologies, we can now define the Bellman equation of dynamic programming:

$$L(w, p_w) = p_w \times \sum_{i \in \mathcal{N}_w} D_i(p_w) + \sum_{v \in V_w} \max_{p_v \geq p_w} L(v, p_v).$$

Here, we require p_v to be no less than p_w, because the price of edge $e = (w, v)$, *i.e.*, $p_e = p_v - p_w$, should be non-negative. The Bellman equation above shows that revenue $L(w, p_w)$ comes from two parts: the revenue of users with destinations at node w (the first term) and the maximum revenue of users with destinations at the descendants of node w (the second term), given that the price of route r_w is p_w.

We have to examine every possible price if we directly solve the above dynamic programming, resulting in exponential computational complexity. The crucial observation is that we can borrow the idea from single link case, and only need to consider a polynomial number of discrete prices without losing much performance. Specifically, in the revised Bellman equation, we only examine the candidate discrete prices from the vector $\hat{\boldsymbol{p}}$, which has been defined as earlier in (5). We can write the Bellman equation in discrete price domain as

$$L(w, \hat{p}_w) = \hat{p}_w \times \sum_{i \in \mathcal{N}_w} D_i(\hat{p}_w) + \sum_{v \in V_w} \max_{\hat{p}_v \geq \hat{p}_w} L(v, \hat{p}_v).$$

We then use the backwards dynamic programming recursion to solve for the optimal price in discrete price domain. Finally, we recursively construct the price $\hat{p}_w^*$ for every possible path r_w. With these path prices, we can set the price p_e for edge $e = (w, v)$ as $p_e = \hat{p}_v^* - \hat{p}_w^*$, and the per unit price for user $i \in \mathcal{N}$ as $p_i = \sum_{e \in r_i} p_e$. We have the following lemma for any subproblem $L(w, p_w)$,[3] which allows us to bound the revenue loss due to the restriction of only considering discrete prices.

Lemma 1. *For the subproblem $L(w, p_w)$, we have*

$$L(w, p_w) \leq (1 + \epsilon) \times L(w, \hat{p}_w),$$

where $\hat{p}_w$ is the largest discrete price that is smaller than p_w, i.e., $\hat{p}_w = (1+\epsilon)^k$ *and* $(1+\epsilon)^k \leq p_w \leq (1+\epsilon)^{k+1}$.

[3] We abuse notations, and use $L(w, p_w)$ and $L(w, \hat{p}_w)$ to denote the subproblems at w that use continuous prices and discrete prices, respectively.

Proof. We can prove this theorem by induction, starting at the leaves of the tree.

• For leaf node w, we have the same revenue maximization problem in the single link case: there exists some integer $k \in [0, K]$ such that $(1+\epsilon)^k \leq p_w \leq (1+\epsilon)^{k+1}$, and

$$\begin{aligned} L(w, p_w) &= p_w \sum_{i \in \mathcal{N}_w} D_i(p_w) \leq (1+\epsilon) \times \hat{p}_w \sum_{i \in \mathcal{N}_w} D_i(\hat{p}_w) \\ &= (1+\epsilon) \times L(w, \hat{p}_w). \end{aligned}$$

Therefore, the lemma holds for leaf nodes.

• For internal node w, suppose the lemma holds for node w's children: *i.e.*, $L(v, p_v) \leq (1+\epsilon) \times L(v, \hat{p}_v)$, for all $v \in V_w$, we then have

$$\begin{aligned} & L(w, p_w) = p_w \times \sum_{i \in \mathcal{N}_w} D_i(p_w) + \sum_{v \in V_w} L(v, p_v) \\ \leq\ & (1+\epsilon)\hat{p}_w \sum_{i \in \mathcal{N}_w} D_i(\hat{p}_w) + (1+\epsilon) \sum_{v \in V_w} L(v, \hat{p}_v) \quad (8) \\ \leq\ & (1+\epsilon) \times L(w, \hat{p}_w). \quad (9) \end{aligned}$$

The inequality (8) follows from the result of single link case and the induction hypothesis. In the continuous price domain, the price p_v of node $v \in V_w$ is always no less than the price p_w of her parent w. We have to preserve this property in the discrete price domain. This property indeed holds, because $\hat{p}_v$ (and $\hat{p}_w$) is the largest discrete price that is smaller than p_v (and p_w). Hence, the set of discrete prices $\{\hat{p}_v | v \in V_w\}$ is a feasible solution to the subproblem at node w. By definition, $L(w, \hat{p}_w)$ is the optimal solution for this subproblem in the discrete price domain. Therefore, we can then derive inequality (9), and get the result for internal nodes.

The proof follows from the above discussion. □

We now have the main theorem for the tollbooth case.

Theorem 2. *For the tollbooth problem with a common source, we can achieve an approximation ratio of $(1+\epsilon)$ with time complexity $O(|V|K^2N)$ and information complexity $O(NK)$.*

Proof. Let OPT and ALG denote the revenue achieved by the optimal solution and our pricing scheme, respectively. Let p_v^* and $\hat{p}_v^*$ be the corresponding optimal price and the price selected by our scheme, respectively. By Lemma 1, we have

$$\begin{aligned} OPT &= \sum_{v \in V_{w_r}} L(v, p_v^*) \leq (1+\epsilon) \sum_{v \in V_{w_r}} L(v, \hat{p}_v) \\ &\leq (1+\epsilon) \sum_{v \in V_{w_r}} L(v, \hat{p}_v^*) = (1+\epsilon) \times ALG, \end{aligned}$$

where the second inequality follows from the fact that ALG is the optimal solution in discrete price domain.

Since we only have to consider discrete prices in the Bellman equation, the dynamic programming table has size $O(|V|K)$. In addition, each entry $L(w, \hat{p}_w)$ can be computed in polynomial time $O(KN)$. Thus, the time complexity is $O(|V|K^2N)$. Similar to the single link case, we need $O(NK)$ rounds of interaction to know the demands of users, resulting in the information complexity of $O(NK)$. □

C. General Network Case

We next consider general network case, and derive three results: two item pricing schemes and one uniform pricing scheme, for different situations. Similar to the idea of single link case, our first pricing scheme is to enumerate every possible price profile of all links, and output the one with the maximum revenue as the final result. The candidate prices for each link also come from the vector $\hat{\boldsymbol{p}}$ in (5). Given a possible set of prices $\boldsymbol{P} = \{p_e\}$, each user i would declare her demand $D_i(p_i)$ to the network provider, where $p_i = \sum_{e \in r_i} p_e$. With the demands from users, the network provider can calculate the overall revenue for each price profile. The network provider finally selects the price profile with the maximum revenue. We have the following theorem for this pricing scheme. The proof is similar to that in single link case, and we reserve it in the technical report [24].

Theorem 3. *For the general network case, we can achieve an approximation ratio of* $(1+\epsilon)$ *with exponential computational complexity* $O(NK^{|E|})$.

Considering the high computational complexity of the above scheme, we turn to two other simple and efficient pricing schemes. Instead of enumerating every possible price for all links, a simple item pricing scheme is to set a single price for all links, which we call single link pricing. Following the same principle as before, we examine the revenue of each possible price from $\hat{\boldsymbol{p}}$ in (5), and select the price with the maximum revenue. We show that such single link pricing scheme is computationally efficient, and has a bounded approximation ratio. We can complete the proof by extending the analysis in single link case, and reserve it to the technical report [24].

Theorem 4. *For the general network case, we can achieve an approximation ratio of* $\bar{p}$ *with polynomial computational complexity* $O(NK)$*, where* $\bar{p}$ *is the upper bound of price.*

The above pricing scheme only guarantees a bounded approximation ratio, which may be pretty large in practice, we now further improve the approximation ratio under some reasonable assumptions. We recall that the hardness of item pricing in a general network arises from the price constraints, which requires the price of a user to be equal to the total price of the links in her selected route. However, other pricing schemes are possible to ensure envy-freeness and arbitrage-freeness. If our goal is to maximize revenue, other pricing schemes may be better and lead to better revenue guarantees. For example, one can first determine link prices as an intermediate step and charge a user something other than the sum of link prices. Therefore, we now generalize our requirement on the pricing scheme, which we present as Assumption 2 below, and discuss why such a pricing scheme is quite reasonable.

Assumption 2. *The per unit price* p_i *to user* i *satisfies a relaxed price constraint:* i.e., $\max_{e \in r_i} p_e \le p_i \le \sum_{e \in r_i} p_e$.

The above assumption on prices is motivated by the requirements of envy-freeness and arbitrage-freeness in practice. On the one hand, as the user with the route over multiple links consumes more resources, the charge to her should be no less than that of the user who uses only one single link, guaranteeing the fairness to some extent. On the other hand, the charge to a user should be no more than the total price of the links in her selected route; otherwise, the user may have an incentive to engage in the following arbitrage behavior: declaring a sequence of data transfer requests, each over a single link, rather than a single request over a route.[4]

With such a price assumption, we can reformulate the problem of revenue maximization as

Problem: *General Revenue Maximization*
Objective: Maximize $\quad L(\boldsymbol{P}) = \sum_{i \in \mathcal{N}} p_i \times x_i$
Subject to:

$$\max_{e \in r_i} p_e \le p_i \le \sum_{e \in r_i} p_e, \qquad \forall i \in \mathcal{N},$$
$$1 \le p_i \le \bar{p}, \qquad \forall i \in \mathcal{N}.$$

We note that the optimal solution to this relaxed optimization is an upper bound of the original revenue maximization problem. Unfortunately, it is still challenging to directly solve the above problem. Here, we are interested in the loss of revenue incurred by using simple envy-free and arbitrage-free pricing schemes. We adopt another simple pricing strategy: charging each user with the same per unit price. We call such pricing scheme as uniform pricing, which has been widely adopted today by Internet service providers and cloud bandwidth providers. We can verify that uniform pricing satisfies Assumption 2, and is envy-free and arbitrage-free.

In order to determine the uniform price and derive a good approximation ratio, we introduce an assumption on the demand functions of users. Such assumption would hold in practice when users have similar utility functions.

Assumption 3. *There exists a representative demand function* $D(p)$*, such that* $\underline{\beta} D(p) \le D_i(p) \le \bar{\beta} D(p)$*, for any user* $i \in \mathcal{N}$ *and any possible price* $p \in [1, \bar{p}]$.

With this assumption, the uniform pricing scheme is quite simple. We optimally solve the optimization problem: Maximize$_{1 \le p \le \bar{p}}$ $p \times D(p)$, and denote the resulting price as $\hat{p}_0^*$.[5] We charge users with the per unit price $\hat{p}_0^*$, and collect revenue $\hat{p}_0^* \times \sum_{i \in \mathcal{N}} D_i(\hat{p}_0^*)$. We show that such uniform pricing scheme can achieve a good approximation ratio.

Theorem 5. *For the general network under Assumptions 2 and 3, we can achieve an approximation ratio of* $\frac{\bar{\beta}}{\underline{\beta}}$ *with the computational complexity of solving Maximize*$_{1 \le p \le \bar{p}}$ $pD(p)$.

Proof. Let $\{p_i^*, i \in \mathcal{N}\}$ to be the optimal prices in the problem of general revenue maximization. According to Assumption 3,

[4]We can further extend Assumption 2 to satisfy more general concept of fairness and to avoid more complex arbitrage behaviors. For example, we can require the unit price p_i of user i to be no less than the price of any sub-path of route r_i. We could also constrain p_i to be no larger than the total price of sub-pathes of r_i, that also form a path between the source and destination.

[5]For some specific demand functions, we can derive the (almost) optimal solution using efficient algorithms, such as applying Golden-section search [25] for strictly unimodular functions. For the general demand function, we can adopt the technique in single link case to obtain the sub-optimal solution with an approximation ratio of $1 + \epsilon$.

we can have

$$
\begin{aligned}
OPT &= \sum_{i\in\mathcal{N}} p_i^* D_i(p_i^*) \le \bar{\beta} \sum_{i\in\mathcal{N}} p_i^* D(p_i^*) \\
&\le \bar{\beta} \sum_{i\in\mathcal{N}} \hat{p}_0^* D(\hat{p}_0^*) \le \frac{\bar{\beta}}{\underline{\beta}} \sum_{i\in\mathcal{N}} \hat{p}_0^* D_i(\hat{p}_0^*). \quad (10)
\end{aligned}
$$

The second inequality follows from that $\hat{p}_0^*$ is the optimal solution of maximizing $pD(p)$. □

It is worth noting that the approximation ratio of uniform pricing only depends on the similarity of demand functions, but not on network topology or data transfer routes.

The remaining task is to construct a representative demand function $D(p)$ that satisfies Assumption 3 and minimizes the ratio $\bar{\beta}/\underline{\beta}$. For general demand functions, it might be difficult to derive the optimal $D(p)$. In the technical report [24], we provide one example using a standard demand function to shed light on the principle of finding $D(p)$. However, this approach requires the knowledge of demand functions. We can achieve an approximation ratio of $\frac{\bar{\beta}}{\underline{\beta}}(1+\epsilon)$, without knowing $D(p)$, by solving the revenue maximization in single link case (4), and regarding the resulting price $\hat{p}^*$ as the uniform price. Using Theorem 1, we can further derive (10) to

$$
(10) \le \frac{\bar{\beta}}{\underline{\beta}} \sum_{i\in\mathcal{N}} p^* D_i(p^*) \le \frac{\bar{\beta}}{\underline{\beta}}(1+\epsilon) \sum_{i\in\mathcal{N}} \hat{p}^* D_i(\hat{p}^*),
$$

where p^* is the optimal price in single link case. We have the following result for this approach.

Theorem 6. *For the general network under Assumptions 2 and 3, we can achieve an approximation ratio of $\frac{\bar{\beta}}{\underline{\beta}}(1+\epsilon)$ with time complexity and information complexity $O(NK)$.*

We trade-off a small revenue loss (an additional factor of $(1+\epsilon)$) to bypass the difficulty in finding the exact representative demand function, which would be more attractive in the scenarios with complicated demand functions.

V. Extension to Multiple Time Slots

We now return to the original problem with multiple time slots, and extend the previous pricing schemes to enable time-dependent pricing [26]. By a slight abuse of notations, we will use the same notations as in the previous section, when appropriate. Due to space limitations, we only express the solution in single link case, which provides insight into the extensions to multiple time slots, and reserve the discussion of tollbooth case and network case to the technical report [24].

The basic idea to design time-dependent pricing scheme is to enumerate possible prices for link(s) at different time slots, and reduce the time complexity via dynamic programming. Due to the assumption that the lengths of time intervals are upper bounded by a constant $\bar{t}$, we just need to examine the prices at only a constant number of time slots in the dynamic programming. Let vector $\boldsymbol{p}_t^{t-\bar{t}+1} = (p_t, p_{t-1}, \cdots, p_{t-\bar{t}+1})$ denote the possible prices for the single link at $\bar{t}$ sequential time slots $t, t-1, \cdots, t-\bar{t}+1$. In the dynamic programming, we maintain a state table $\tilde{L}(t, \boldsymbol{p}_t^{t-\bar{t}+1})$, where each entry records the maximum revenue that we can obtain from users with time intervals ending at time slot t or before, given that the prices at time slots $t, t-1, \cdots, t-\bar{t}+1$ are set as $p_t, p_{t-1}, \cdots, p_{t-\bar{t}+1}$, respectively. We use $\mathcal{N}_t$ to denote the set of users, whose time intervals ends exactly at slot t. We now can define the Bellman equation for dynamic programming

$$
\tilde{L}(t, \boldsymbol{p}_t^{t-\bar{t}+1}) = \sum_{i\in\mathcal{N}_t} p_i D_i(p_i) + \max_{p_{t-\bar{t}}} \tilde{L}(t-1, \boldsymbol{p}_{t-1}^{t-\bar{t}+1}, p_{t-\bar{t}}),
$$

where p_i is the per unit price to user i, *i.e.*, $p_i = \sum_{t\in[t_i^1, t_i^2]} p_t$. We note that price vector $\boldsymbol{p}_t^{t-\bar{t}+1}$ is sufficient to calculate the price p_i and then the revenue of users $\mathcal{N}_t$ (the first term of the equation), because no time interval has length larger than $\bar{t}$. The second term of the equation is the maximum revenue among the subproblem at time slot $t-1$ with prices $\boldsymbol{p}_{t-1}^{t-\bar{t}+1}$ to time slots $t-1, t-2, \cdots, t-\bar{t}+1$. Similar to the development in the previous section, we only consider discrete prices in the Bellman equation. Using the standard backward recursion approach, we can solve the dynamic programming in discrete price regime, and obtain the price $\hat{p}_t$ for each time slot t. The following lemma bounds the revenue loss due to only considering discrete prices in the dynamic programming.

Lemma 2. *For any state $\tilde{L}(t, \boldsymbol{p}_t^{t-\bar{t}+1})$, there always exists a state $\tilde{L}(t, \hat{\boldsymbol{p}}_t^{t-\bar{t}+1})$ using only discrete prices, such that*

$$
\tilde{L}(t, \boldsymbol{p}_t^{t-\bar{t}+1}) \le (1+\epsilon) \times \tilde{L}(t, \hat{\boldsymbol{p}}_t^{t-\bar{t}+1}),
$$

where $\hat{\boldsymbol{p}}_t^{t-\bar{t}+1} = (\hat{p}_t, \hat{p}_{t-1}, \cdots, \hat{p}_{t-\bar{t}+1})$, and $\hat{p}_t$ is the largest discrete price that is smaller than p_t, i.e., *$\hat{p}_t = (1+\epsilon)^k$ and $(1+\epsilon)^k \le p_t \le (1+\epsilon)^{k+1}$ for some integer $k \in [0, K]$.*

Using the preceding result, we can show the performance guarantee of the time-dependent pricing for single link case.

Theorem 7. *For the single link with multiple time slots, we can achieve an approximation ratio of $(1+\epsilon)$ with time complexity $O(TK^{\bar{t}+1}N)$ and information complexity $O(NK)$, where T is the number of time slots.*

Due to space limitations, we put the detailed proofs of Lemma 2 and Theorem 7 into our technical report [24].

VI. Evaluation Results

In this section, we only provide evaluation results for the general network case, considering that we can obtain near-optimal results for the single link and tollbooth network cases. We also observe that the evaluation result in multiple time slot setting is similar to that in single time slot, and thus we only report results for the special single time slot case.

A. Methodology

We first present the evaluation setting. The network topology we consider in our evaluation is B4, a WAN connecting Google's data centers across the world in 2011 [2]. In this WAN graph, there are 19 links/edges connecting 12 regions/nodes. We randomly generate a pair of source and destination nodes for each user, and regard the shortest path between the source and destination as her selected route to

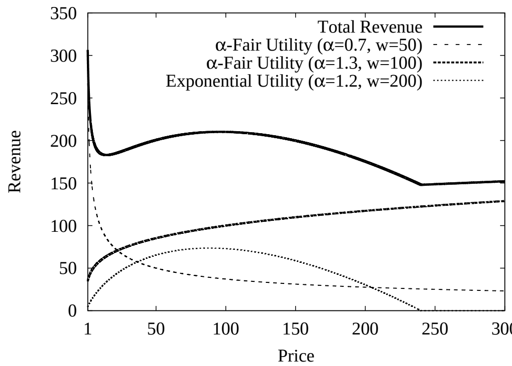


Fig. 2. Revenue curves for two different α-fair utility functions (one with parameters $\alpha = 0.7$, $w = 50$ and the other with $\alpha = 1.3$, $w = 100$) and one exponential utility function with parameters $\alpha = 1.2$, $w = 200$.

transmit data. We consider two specific classes of utility functions: α-fair utility function

$$u_i(x_i) = \begin{cases} w_i \frac{x_i^{1-\alpha_i}}{1-\alpha_i}, & \alpha_i > 0, \alpha_i \neq 1, \\ w_i \log x_i, & \alpha_i = 1, \end{cases}$$

and exponential utility function $u_i(x_i) = w_i(1 - e^{-\alpha_i x})$. By the definition of demand function in (3), we can obtain the corresponding revenue functions

$$L_i(p) = p \times \left(\frac{p}{w_i}\right)^{-\frac{1}{\alpha_i}} \text{ and } \quad L_i(p) = -\frac{p}{\alpha_i} \times \log \frac{p}{w_i \alpha_i}$$

for α-fair utility and exponential utility, respectively. In Figure 2, we plot the curves for three revenue functions with different parameters over price range $[1, 300]$. From the figure, we can observe that for $\alpha > 1$, the revenue of α-fair utility increases with price p, whereas for $\alpha < 1$ it decreases with price. The revenue function of exponential utility is concave, reaches the maximum value at price $w_i\alpha_i/e$, and becomes negative when price is larger than $w_i\alpha_i$. In Figure 2, we also draw the overall revenue function, which is non-convex and demonstrates the difficulty in directly solving the revenue maximization problem. Each of users is associated with either an α-fair utility or an exponential utility with equal probability. The parameters (α_i and w_i) of utility functions are draw from some distributions. Specifically, α_i follows a uniform distribution over the interval $[0.5, 1.5]$, and w_i follows either a uniform distribution or a Pareto distribution. We adopt Pareto distribution for w_i to model the scenario that most of users have similar utilities. As we observe in Figure 2, the revenue of exponential utility would be negative when $p > w_i\alpha_i$, and thus we set a relatively large w_i for exponential utility. The supports of the uniform distributions for w in α-fair utility and exponential utility are set as $[1, 60]$ and $[1, 1000]$, respectively. The scale and shape pairs of the Pareto distributions for w in α-fair utility and exponential utility are set as $(1, 1)$ and $(50, 1)$, respectively. We investigate the performance of pricing schemes with different number of users, varying from 20 to 200 with increment of 20, and with different upper bounds of prices, ranging from 20 to 200 with increase 20. All the evaluation results are averaged over 1000 runs.[6]

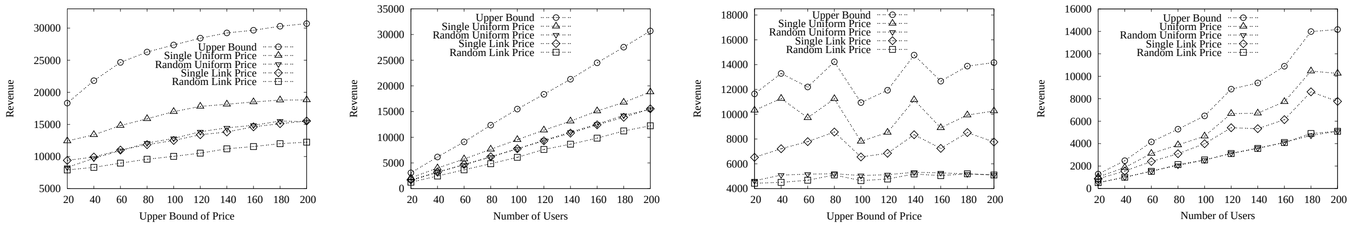
We implement uniform pricing and single link pricing, and compare their performance with that of random uniform pricing and random single link pricing. Considering that we cannot optimally solve the general revenue maximization in Section IV-C, we compute a trivial upper bound of the optimal revenue by solving Maximize$_{1 \leq p_i \leq \bar{p}}$ $\sum_{i \in \mathcal{N}} p_i \times D_i(p_i)$. As we do not have specific price correlation among users in this optimization, we can separately maximize each term in the objective, which is easy to deal with in the context of α-fair utility and exponential utility. We compare the performance of pricing schemes with this upper bound, to measure the revenue loss of considering the envy-freeness and arbitrage-freeness.

B. *Performance of Pricing Schemes*

Figure 3 shows the revenue achieved by different pricing schemes under various evaluation settings. Generally, we can see from Figure 3(b) and Figure 3(d) that in the market with larger number of users, all pricing schemes obtain higher revenue. For the cases of w-Uniform distribution (Figure 3(a)), the revenue increases with $\bar{p}$, because pricing schemes have higher probability to obtain large revenue when there are more candidate prices to choose. However, when w follows a Pareto distribution (Figure 3(c)), the revenue does not have an explicit relation with $\bar{p}$. Although Pareto distribution only has small probability to generate a large w, if this happens (independent of $\bar{p}$), the revenue from the user with such large w would dominate the revenue of other users, leading to a surge in the overall revenue.

We now compare the performance of uniform pricing with that of other pricing schemes to show its advantage in maximizing revenue. From Figure 3, we can observe that the ratio between the upper bound of optimal revenue and the revenue achieved by uniform pricing is not so large, especially when w follows a Pareto distribution. This indicates that we do not sacrifice too much revenue to guarantee the properties of envy-freeness and arbitrage-freeness. Theorem 5 states that the approximation ratio of uniform pricing is $\bar{\beta}/\underline{\beta}$, which is related to the similarity of utility (or demand) functions among users. The observations from the evaluation results explicitly indicate this relation. As shown in Figure 3(a), the ratio stays at around 1.6 for all the possible $\bar{p}$. The reason is that the similarity of utility functions does not change much in different rounds when N is fixed. In contrast, we can see from Figure 3(b) that the ratio becomes large with the increase of N, because utility functions would be more diverse when N is larger. Again, we have the same observations from Figure 3(c) and Figure 3(d) when w follows a Pareto distribution. We also find that the ratio in the case of w-Pareto Distribution is smaller than that in the case of w-Uniform Distribution under the same simulation setting. This is because most of w's generated by Pareto distribution locate densely in a small interval, reflecting the high similarity of utility functions.

[6]All parameters can be different from the ones used here. As the evaluation results of using other parameters are similar, we only report the results for these parameters in this paper.



(a) w-Uniform Distribution, $N = 200$ (b) w-Uniform Distribution, $\bar{p} = 200$ (c) w-Pareto Distribution, $N = 200$ (d) w-Pareto Distribution, $\bar{p} = 200$

Fig. 3. Performance of different pricing schemes under various evaluation settings.

From Figure 3, we can observe that uniform pricing always outperforms single link pricing, which demonstrates that we can extract more revenue by relaxing the price constraints in item pricing, and using the price constraints in Assumption 2. In single link pricing scheme, we require that the per unit price to a user should be proportional to the length of her route. As we also require the per unit price to be less than an upper bound, the feasible prices to links would be quite small if there exist some users who transmit data over long routes in network. This would result in large revenue loss especially when the overall revenue increases with price.

In Figure 3, both uniform pricing and single link pricing achieve higher revenue than their random counterparts. This indicates that we can significantly improve the performance of pricing schemes through a polynomial number of interactions with users. These interactions provide information for the network provider to optimize price and then extract more revenue. Thus, the network provider has an economic incentive to interact with users to learn their demands,

VII. Conclusion and Future Work

In this paper, we have studied a revenue maximization problem in inter-datacenter networks, where a network service provider sets the price to each user and then users decide a certain rate to transmit data over the network. We focus on the family of envy-free and arbitrage-free pricing schemes. For single link and tollbooth cases, we have designed a $(1+\epsilon)$-approximation pricing scheme with polynomial time complexity and information complexity. For general network case, we have established the following result: when users have similar utility functions, uniform pricing can achieve a good approximation ratio, which is independent of network topology and data transfer requests.

One possible direction for future work is to consider capacity constraints and congest effect in pricing scheme design. Another interesting research topic is to design item pricing for revenue maximization in general network topology.

References

[1] V. K. Adhikari, Y. Guo, F. Hao, M. Varvello, V. Hilt, M. Steiner, and Z. L. Zhang, "Unreeling netflix: Understanding and improving multi-cdn movie delivery," in *INFOCOM*, 2012.

[2] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, J. Zolla, U. Hölzle, S. Stuart, and A. Vahdat, "B4: Experience with a globally-deployed software defined wan," in *SIGCOMM*, 2013.

[3] C.-Y. Hong, S. Kandula, R. Mahajan, M. Zhang, V. Gill, M. Nanduri, and R. Wattenhofer, "Achieving high utilization with software-driven wan," in *SIGCOMM*, 2013.

[4] "AWS Direct Connect:," https://aws.amazon.com/directconnect/.

[5] V. Jalaparti, I. Bliznets, S. Kandula, B. Lucier, and I. Menache, "Dynamic pricing and traffic engineering for timely inter-datacenter transfers," in *SIGCOMM*, 2016.

[6] V. Guruswami, J. D. Hartline, A. R. Karlin, D. Kempe, C. Kenyon, and F. McSherry, "On profit-maximizing envy-free pricing," in *SODA*, 2005.

[7] T. Başar and R. Srikant, "A stackelberg network game with a large number of followers," *Journal of Optimization Theory and Applications*, vol. 115, no. 3, pp. 479–490, 2002.

[8] ——, "Revenue-maximizing pricing and capacity expansion in a many-users regime," in *INFOCOM*, 2002.

[9] S. Shakkottai, R. Srikant, A. Ozdaglar, and D. Acemoglu, "The price of simplicity," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1269–1276, 2008.

[10] P. Marbach and R. Berry, "Downlink resource allocation and pricing for wireless networks," in *INFOCOM*, 2002.

[11] S. Li and J. Huang, "Price differentiation for communication networks," *IEEE/ACM Transactions on Networking*, vol. 22, no. 3, pp. 703–716, 2014.

[12] P. Hande, M. Chiang, R. Calderbank, and J. Zhang, "Pricing under constraints in access networks: Revenue maximization and congestion management," in *INFOCOM*, 2010.

[13] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, Mar 1998.

[14] R. Srikant, *The mathematics of Internet congestion control*. Cambridge, MA: Birkhauser, 2004.

[15] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, 2007.

[16] D. Acemoglu, A. Ozdaglar, and R. Srikant, "The marginal user principle for resource allocation in wireless networks," in *CDC*, 2004.

[17] M.-F. Balcan, A. Blum, and Y. Mansour, "Item pricing for revenue maximization," in *EC*, 2008.

[18] R. B. Myerson, "Optimal auction design," *Mathematics of operations research*, vol. 6, no. 1, pp. 58–73, 1981.

[19] S. Kandula, I. Menache, R. Schwartz, and S. R. Babbula, "Calendaring for wide area networks," in *SIGCOMM*, 2014.

[20] A. Kumar, S. Jain, U. Naik, A. Raghuraman, N. Kasinadhuni, E. C. Zermeno, C. S. Gunn, J. Ai, B. Carlin, M. Amarandei-Stavila, M. Robin, A. Siganporia, S. Stuart, and A. Vahdat, "BwE: Flexible, hierarchical bandwidth allocation for wan distributed computing," in *SIGCOMM*, 2015.

[21] T. Roughgarden and E. Tardos, "How bad is selfish routing?" *Journal of the ACM*, vol. 49, no. 2, pp. 236–259, 2002.

[22] J. F. Traub, *Information-based complexity*. John Wiley and Sons Ltd., 2003.

[23] R. Johari and J. N. Tsitsiklis, "Efficiency of scalar-parameterized mechanisms," *Operations Research*, vol. 57, no. 4, pp. 823–839, 2009.

[24] "Pricing for revenue maximization in inter-datacenter networks," Tech. Rep., 2017, available at https://drive.google.com/open?id=0BzPQ3WpemY1VTzFjNk1kbzF3QzA.

[25] D. P. Bertsekas, *Nonlinear programming*. Athena scientific Belmont, 1999.

[26] L. Jiang, S. Parekh, and J. Walrand, "Time-dependent network pricing and bandwidth trading," in *IEEE Network Operations and Management Symposium Workshops*, 2008.