

面向数据中心间网络带宽的在线定价机制设计： 基于强化学习的方法

牛超越 陈培煜 张嘉懿 吴 帆 陈贵海

(上海交通大学计算机科学与工程系 上海 200240)

摘 要 随着云服务的快速发展,数据中心间的网络带宽已经成为了宝贵的资源。目前,带宽以固定价格出售,并由流量工程进行容量分配。这种方式无法提供传输时间的保证。然而,大多数传输任务都有严格的截止时间限制,未按时完成传输会给用户造成较大的收益损失。此外,对于服务提供商来说,固定价格定价机制将导致网络利用率低下,不利于最大化其长期累积收益。因此,本文希望设计在线定价机制,最大化服务提供商的收益,同时保证用户传输请求的真实性。本文首先提出基于价格表的在线定价机制 PLPM:服务提供商向用户实时地展示价格表,由用户选择特定的传输时间和传输量以满足其传输要求。PLPM 利用强化学习的方法更新价格表,实现对不同时间、不同容量状态的带宽进行定价。进一步地,本文提出了基于请求的在线定价机制 RPM:用户可以自定义其传输类型、传输量和截止时长。RPM 在线地学习用户的估价以实现收益最大化和真实性,并匹配了基于优先级的带宽分配策略以提升网络带宽利用率。最后,通过实验证明了所提出的定价机制相比于固定定价机制,可以大幅提高累积收益和网络带宽的利用率。

关键词 在线定价;数据中心间网络带宽;强化学习;收益最大化;真实性

中图法分类号 TP393

DOI号 10.11897/SP.J.1016.2022.01068

Reinforcement Learning-Based Online Pricing for Inter-Datacenter Bandwidth

NIU Chao-Yue CHEN Pei-Yu ZHANG Jia-Yi WU Fan CHEN Gui-Hai

(Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200240)

Abstract With the rapid development of cloud-based services, bandwidth between datacenters has become a valuable resource. Currently, the bandwidth is sold with a fixed price and is allocated using traffic engineering mechanisms, which have no guarantee on when the transfers can be finished. However, most transfer tasks have strict deadlines, while not finishing them on time may result in high losses to customers. In addition, trivially fixed pricing strategies will cause network inefficiency and cannot maximize service providers' revenues. We thus focus on how to design online pricing mechanisms for inter-datacenter transfers, while achieving revenue maximization and truthfulness. We first propose Price List-Based Online Pricing Mechanism (PLPM), providing customers with a price list to select specific time slots and amounts. In particular, discriminative prices for bandwidth at different time slots with corresponding capacities are calculated by reinforcement learning methods. We then extend PLPM to Request-Based Online Pricing Mechanism (RPM), which allows customers to request customized data transfers. RPM online explores and exploits the valuations of customers for revenue maximization and meanwhile

收稿日期:2020-12-22;在线发布日期:2021-08-19。本研究得到科技创新 2030—“新一代人工智能”重大项目(2018AAA0100900)、国家自然科学基金项目(62025204,61972254)、阿里巴巴创新研究计划资助。牛超越,博士,主要研究方向为数据共享与交易、移动计算、机器学习、隐私安全。E-mail: rvince@sjtu.edu.cn。陈培煜,硕士研究生,主要研究方向为数据定价、网络经济学。张嘉懿,硕士,主要研究方向为网络经济学、强化学习。吴 帆(通信作者),博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为网络经济学、无线网络、移动计算、隐私安全。E-mail: fwu@cs.sjtu.edu.cn。陈贵海,博士,教授,中国计算机学会(CCF)会士,主要研究领域为分布式计算、计算机网络。

guarantees truthfulness against rational customers. Allocation rules with priorities are designed to improve network utilization. We finally conduct experiments and demonstrate that the proposed mechanisms outperform fixed pricing from cumulative revenue and network utilization.

Keywords online pricing; inter-datacenter bandwidth; reinforcement learning; revenue maximization; truthfulness

1 引言

在过去几年里,全球对于云服务的需求不断增长,用于连接不同数据中心的广域网带宽资源越发宝贵^[1-2].例如,搜索应用需要高通量的网络传输来更新和传送新索引^[3];视频交付服务需要低延迟的网络传输来提升用户体验^[4];大数据应用需要将数据集从一个数据中心传输到另一个数据中心进行数据融合、数据挖掘、机器学习和深度学习等任务^[5].国内外许多云平台(例如国内的阿里云、华为云和腾讯云,美国的亚马逊 AWS、谷歌云和微软 Azure 等)已经在全球建立了数据中心和用于数据传输服务的广域网.

本文主要从服务提供商的角度出发考虑数据中心间网络带宽的在线定价机制设计.目前,服务提供商针对具有相同起点和终点的传输任务采用每单位带宽相同价格,例如表 1 所示的谷歌高端分级定价策略.然而,固定定价机制主要存在以下 3 点弊端:(1)根据不同的数据传输需求,用户的任务请求可分为以下 3 种类型^[6-7]:①固定速率:需要在特定的时间内分配固定的带宽;②固定期限:需要在特定的时间内完成传输任务;③后台传输:没有任何截止时间的限制.显然地,固定定价机制无法满足固定期限传输的需求.根据微软的调研^[8]发现:多数传输任务都有严格的截止时间要求,如果超过了特定时间,用户将会遭受巨额损失.同时,大部分用户愿意支付更多的费用以保证自身的数据传输需求.因此,服务提供商如果能够在满足用户个性化传输需求的前提下收取更高的价格,则可以提高其收益;(2)固定定价会导致网络带宽利用率低下并降低服务提供商的收益.由于没有价格差异,用户可以随时提交传输请求.这将导致网络链路在一定时间段内处于高峰状态,有较高的拥塞可能性,而在另一段时间内则完全空闲,带宽利用率低.网络拥塞和网络空闲的情况都将严重损害服务提供商的收益;(3)固定定价不能使服务提供商在动态的市场环境下实现收益的

最大化.服务提供商如果能够观察学习动态的网络带宽需求状况并合理地调整价格,将比采用朴素的固定定价机制获得更高的收益.

表 1 谷歌广域网带宽的高端分级定价表

传输起点	传输终点	单位价格(美元/GB)
北美	北美	0.08~0.12(0~10 TB)
		0.08(10~150 TB)
		0.06(150~500 TB)
亚洲	亚洲	0.08~0.12(0~10 TB)
		0.085(10~150 TB)
		0.08(150~500 TB)
大洲 1	大洲 2	0.08~0.12

为了规避上述固定定价机制的弊端,需要为数据中心间的网络传输设计一个实用、高效、鲁棒的在线定价机制,并重点解决以下 3 个挑战问题:(1)如何构建合适的市场模型?为了设计以收益最大化为目标的定价机制,需要合理地建模服务提供商和用户之间复杂的供需关系.其中,当前网络带宽容量状态、传输类型、待传输数据的总量、传输截止时间等多个因素会影响用户的估价.现在的相关工作^[5,8]尚未考虑这些因素;(2)如何实时地学习用户对数据传输的估价?目前,在无估价模型的定价机制设计方面,已有适应市场动态变化、基于强化学习的解决方案^[9-14].这些研究工作为了降低计算复杂度直接把待传输的数据量作为学习的状态,亦或采用拍卖的形式.然而,一方面,数据中心间的传输必须将上述因素作为状态包括在内,而计算时间却随着状态变量的增长而呈指数增长;另一方面,拍卖以及分配产生的时延对于需要进行紧急传输任务的用户来说是无法接受的;(3)如何保证用户传输请求的真实性?在现实中,理性且自私的用户可能会谎报他们的传输请求以降低开销、提高收益,例如谎报传输的数据量、传输的截止时间或多种谎报策略的组合等.理想的定价机制需要在保证用户请求真实性的情况下实现收益最大化的目标.

本文提出了两种基于强化学习的在线带宽定价机制:PLPM(Price List-based online Pricing Mechanism)和 RPM(Request-based online Pricing

Mechanism), 以实现服务提供商收益的最大化和用户请求的真实性. 首先, 在 PLPM 中, 服务提供商制定一个价格表, 包含每个时间的价格和剩余带宽容量. 用户根据他们的需求和估价选择传输时间和传输量. 价格表通过无模型的 Q 学习算法(Q-learning)进行实时更新, 并将剩余容量作为一个状态来计算价格, 即 Q 学习得出的动作, 以实现服务提供商的收益最大化. 在 PLPM 的基础上, 本文提出支持用户细粒度传输请求的 RPM. 用户可以提交他们个性化的请求, 包括传输类型、传输量和传输截止时间, 而 RPM 会对用户的估价进行探索. RPM 通过对定价函数进行了一些约束以保证用户请求的真实性. 针对无截止时间的传输, 例如后台传输, RPM 采用无状态学习的方式探索最佳价格, 具体采用了多臂老虎机(Multi-Armed Bandit, MAB)相关的算法. 针对有截止时间限制的传输, RPM 将传输量和截止时间作为定价函数的输入, 并将传输截止时间内的总容量作为 Q 学习中的状态, 以获取候选价格. 在带宽分配机制方面, 本文将后台传输置于低优先级, 并设计了高效的分配算法以提高网络带宽的利用率. PLPM 和 RPM 都使用 Q 学习, 将网络容量作为状态, 并将收益作为奖励, 实时地调整价格. 相比于 PLPM, RPM 考虑了用户请求中不同的传输类型和自定义的截止时间, 并设计了相应的动态价格计算方式和网络带宽分配方案. 下面总结这篇工作的核心贡献:

- (1) 首先, 本文对服务提供商和用户之间动态的网络带宽交易场景进行建模, 并刻画了在线定价机制需要满足的性质, 包括收益最大化和请求真实性.
- (2) 其次, 提出了基于强化学习的 PLPM 和 RPM 在线定价机制: PLPM 根据不同时间、不同的带宽剩余容量制定不同的价格, 并为用户统一提供相应的时间-价格表; RPM 进一步地考虑了不同用户多样的数据传输请求以及高度差异化的估价, 并为不同类型的传输请求设计定价策略. 本文还为在线定价机制匹配了一种基于优先级的带宽分配算法从而提升网络带宽利用率. 相比于该方面已有的研究, 本文首次同时考虑用户请求的具体内容, 实现服务提供商收益最大化, 并且优化带宽利用率.
- (3) 再次, 从理论上证明了 RPM 对于用户请求真实性的保证. 特别地, 本文提出了通过选择定价函数中的参数作为 Q 学习中的动作从而实现真实性的新方法. 据我们所知, 这是将真实性纳入基于强化学习的动态定价机制的首创.

(4) 最后, 实验结果表明: PLRM 和 RPM 优于传统的固定定价机制, 且 RPM 实现了更高的收益率和网络带宽利用率. 特别地, 在算法收敛情况下, RPM 的累积收益达到理想最优累积收益的 80%.

2 云带宽定价模型

本节首先介绍数据中心间网络传输的系统模型. 然后将引入本文所考虑的在线定价问题模型, 并介绍定价机制的理想特性. 最后将介绍相关的背景知识. 本文常用的数学符号及其含义如表 2 所示.

表 2 本文常用的数学符号及其含义

符号	含义
$\mathcal{N}=\{1,2,\cdots,N\}; i \in \mathcal{N}$	带宽市场的用户全集; 编号 i 的用户
$\mathcal{T}=\{1,2,\cdots,T\}; t \in \mathcal{T}$	时间点全集; 第 t 个时间点
c_e	网络链路 e 的最大带宽容量
$c_{e,t}$	网络链路 e 在时间 t 的带宽余量
c'	带宽售卖的最小单位容量
$q_i=(\xi, d_i, t_i^1, \Delta t_i)$	用户 i 真实所需的传输请求, 包含传输的类型、数据量、开始时间、截止时长
BT; FDT	后台传输; 固定期限传输
$\hat{t}$	最大的截止时长, 即 Δt_i 的上界
$Q_i=\{q_{i1}, q_{i2}, \cdots, q_{im}\}$	用户 i 实际提交的传输请求集合
$v_i(\cdot)$	用户 i 的估价函数
$p(\cdot)$	服务提供商的定价函数
$\hat{p}, \hat{p}$	单位数据传输候选价格的上下界
$\bar{p}=\{\bar{p}_1, \bar{p}_2, \cdots, \bar{p}_K\}$	候选价格集合(作为 PLPM 定价机制中 Q 学习算法的动作, 以及 RPM 定价机制中 UCB 算法的动作)
u_i	用户 i 的收益, 即估价和价格之间的差值
$x_i \in \{0,1\}$	用户 i 是否接受服务提供商的价格
$s_t, a_t, Q(s_t, a_t), r_t$	在时间 t 的状态、选择的动作、Q 值、对应的奖励
ϵ	ϵ -greedy 算法的概率参数
n_a	UCB 算法中选取动作 a 的总次数
α, γ	Q 学习算法中的学习率与折扣因子
θ	RPM 定价机制中定价函数的参数并作为 Q 学习算法的动作

2.1 市场模型

网络 本文将数据中心间的网络图表示为 G , 将边集(即所有网络链路)表示为 E , 并将某条网络链路 $e \in E$ 的初始带宽容量表示为 c_e . 实际中的网络带宽容量为连续值, 但在本文所考虑的交易定价场景中存在带宽售卖的最小容量单位. 因此, 带宽容量可离散化处理, 且最大带宽容量 c_e 可表示为最小售卖单位的整数倍. 此外, 将连续时间离散化为不相交的时间点集 $\mathcal{T}=\{1, \cdots, T\}$ ^[15], 其中 $t \in \mathcal{T}$ 代表某个时间. 用 $c_{e,t}$ 表示网络链路 e 在时间 t 尚未分配的带宽余量.

用户 用 $\mathcal{N}=\{1,2,\cdots,N\}$ 表示一组有数据传输需求的用户. 根据文献[6-7], 用户的数据传输任务

可以分为以下三种主要类型.

(1) 后台传输 (Background Transfer, BT). 没有截止时间或截止时间较长, 但该类型请求通常有大量的数据需要传输.

(2) 固定期限传输 (Fixed Deadline Transfer, FDT). 要求在截止时间之前传输一定量的数据. 如果在截止期限之前未完成传输, 用户将产生较大的损失.

(3) 固定速率传输 (Fixed Rate Transfer, FRT). 需要在一定时间内以恒定速率传输数据. 这种类型的传输请求对延迟非常敏感. 例如, 面向移动终端应用程序的数据传输和面向数据中心的高清视频传输, 需要高传输速率以保证服务质量, 即使 100 ms 的时延也会大幅地降低用户体验.

固定速率传输可视为固定期限传输的一种特殊情况, 即将固定速率传输视为在每个时间具有相同传输量并存在固定期限的传输序列. 本文用元组 $q_i = (\xi_i, d_i, t_i^1, t_i^2)$ 来表示用户 $i \in \mathcal{N}$ 真实的传输任务请求. 其中, q_i 的传输类型 $\xi_i \in \{\text{BT}, \text{FDT}\}$. 如果用户 i 请求后台传输 BT, 该用户需要指定传输总量 d_i 和传输的开始时间 t_i^1 ; 如果用户 i 请求固定期限传输 FDT, 除传输总量和开始时间外, 该用户还需要指定截止时间 t_i^2 . 为了简化问题模型, 本文将开始时间 t_i^1 视为用户 i 到达市场的时间, 并用 Δt_i 表示开始时间 t_i^1 和截止日期 t_i^2 之间的时长, 即 $\Delta t_i = t_i^2 - t_i^1$. 同时, 将最大的任务时长表示为 $\hat{t}$. 对于后台传输, $\forall i \in \mathcal{N}, \xi_i = \text{BT}, \hat{t} = \infty$; 对于固定期限传输, $\forall i \in \mathcal{N}, \xi_i = \text{FDT}, \hat{t} \geq \Delta t_i$. 用集合 $Q_i = \{q_{i1}, q_{i2}, \dots, q_{im}\}$ 表示用户 i 实际发起的系列性请求. 当用户 i 真实地提交传输请求时, 实际请求和真实请求相同, 即 $Q_i = \{q_i\}$. 然而, 在现实中, 理性自私的用户可能提交虚假请求以降低价格.

估价 每个用户对能满足传输截止时间的请求有私人的估价, 即用户对该任务愿意付出的最高价格. 本文主要针对具有严格截止时间约束的固定期限传输请求, 例如不可分割的数据或者接收方在无法及时获得完整数据将产生巨大损失的应用^[16]. 换句话说, 如果错过了截止时间, 则传输是无用的, 用户也将无法获得任何收益. 本文对估价函数作了以下定义.

定义 1 (估价函数). 用户 i 对于其传输任务 q_i 的估价函数 $v_i(\cdot)$ 需满足: (1) 在 Δt_i 时间内完成 d_i 量的数据传输, 用户 i 才能获得收益; (2) 用户 i 无法通过提前完成任务或请求更多的带宽 (例如在 $\Delta t \leq$

Δt_i 时间内完成 $d \geq d_i$ 量的数据传输) 提高收益. 形式化的表达如下:

$$v_i(\xi_i, d, \Delta t) = \begin{cases} 0, & d < d_i \text{ or } \Delta t > \Delta t_i \\ v_i(\xi_i, d_i, \Delta t_i), & d \geq d_i \text{ and } \Delta t \leq \Delta t_i \end{cases} \quad (1)$$

价格 服务提供商对用户的传输请求设定价格. 用定价函数 $p(q(\xi, d, \Delta t, C))$ 表示服务提供商根据用户请求在 Δt 时间内传输 d 单位数据收取的费用, 其中 C 表示所有时段的可用带宽矩阵. 为了简化表示, 用 $p(d; \xi_i, \Delta t_i, C)$ 表示在给定 $\xi_i, \Delta t_i, C$ 的情况下传输 d 单位数据的价格, 用 $p(\Delta t; \xi_i, d_i, C)$ 来表示在给定 ξ_i, d_i, C 的情况下 Δt 时间内完成数据传输任务的价格. 本文假设价格为非负数, 单位数据传输候选价格的范围为 $[\bar{p}, \hat{p}]$.

收益 看到服务提供商给出的价格后, 理性自私的用户 (以自己的收益最大化为目标而不关注网络带宽的利用率等其他因素) 将根据其收益来决定是否传输数据. 用户 i 的系列请求 Q_i 带来的收益可以表示为用户的估价与服务提供商的出价之间的差值, 即

$$u_i = v_i(Q_i) - \sum_{q_{ij} \in Q_i} p_i(q_{ij}, C) \quad (2)$$

为了简化表示, 用 $u(d; \xi_i, d_i, \Delta t_i, C)$ 表示给定 $\xi_i, d_i, \Delta t_i, C$ 下传输 d 单位数据的收益, 用 $u(\Delta t; \xi_i, d_i, \Delta t_i, C)$ 表示给定 $\xi_i, d_i, \Delta t_i, C$ 下在 Δt 时间内完成数据传输任务的收益. 因此, 如果用户的收益大于等于零, 则用户将接受服务提供商的价格, 发起传输请求; 否则, 用户将不会购买传输服务. 用 x_i 表示用户 i 是否接受价格, 即

$$x_i = \begin{cases} 0, & u_i < 0 \\ 1, & u_i \geq 0 \end{cases} \quad (3)$$

2.2 交易模型

图 1 展示了数据中心间网络带宽市场的交易模型. 用户首先发起数据传输请求, 服务提供商根据当前的网络状态 (例如每个时间的带宽剩余容量) 和用户的请求内容 (包括传输类型、截止时间和传输数据总量等) 制定价格. 获得价格后, 用户将根据自己估价及收益, 选择是否付费传输数据. 如果用户接受服务提供商的价格并付费, 服务提供商将安排执行用户的传输任务. 具体来说, 服务提供商的带宽容量调度模块将从网络中获取每个时间的当前剩余带宽, 然后为用户的请求分配带宽以保证任务的完成, 同时使网络带宽有较高的利用率. 任务分配完成后, 网络带宽的剩余容量将发生变化, 网络状态也将相应地更新. 为简化交易模型, 本文考虑不同用户提交传

输请求的时间戳唯一,且服务提供商可以对用户的在线请求进行统筹安排.例如,针对同一时间不同用户提交相同请求的特殊情况,服务提供商可以随机将用户请求进行排序.此外,本文假设用户决策是否接受价格的时间可忽略.

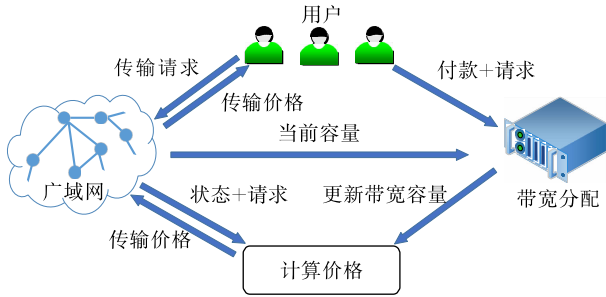


图 1 数据中心间网络带宽交易流程图

针对上述市场模型,本文考虑两种定价机制:价格表模式的 PLPM(第 3 节)和支持自定义传输请求的 RPM(第 4 节).这两种机制在提交请求、计算价格和带宽分配等方面有所不同. RPM 是 PLPM 的拓展,价格的计算考虑了请求的具体内容,而且能根据用户是否拒绝服务提供商的出价在线学习用户的估价函数.

2.3 问题模型

本文中定价机制设计的目标主要包括:(1)服务提供商的累积收益最大化;(2)用户提交请求的真实性保证.

累积收益最大化 面向一组用户的传输请求 $\{Q_i | \forall i \in \mathcal{N}\}$, 最大化服务提供商累积收益的目标可表示为

$$\begin{aligned} \max \quad & R = \sum_{i \in \mathcal{N}} \sum_{q_{ij} \in Q_i} p(q_{ij}) \times x_i, \\ \text{s. t.} \quad & \forall i \in \mathcal{N}, q_{ij} \in Q_i, \tilde{p} \leq \frac{p(q_{ij})}{d_{ij}} \leq \hat{p}, x_i \in \{0, 1\} \end{aligned} \quad (4)$$

其中 q_{ij} 表示用户 i 第 j 个传输请求, d_{ij} 表示 q_{ij} 的传输数据量.

用户请求真实性 在价格表模式下,用户提交请求前就能看到(后续)每个时间网络带宽的单位价格.用户将根据自己的传输量在传输截止时间前尽可能选择便宜的时间进行数据传输.鉴于最便宜方案的唯一性,用户只会真实地选择每个时段所要传输的数据量,而用户的选择与用户的真实估价无关.因此,PLPM 的设计无需考虑真实性.然而,在用户自定义请求模式下,服务提供商的价格是在用户提交请求后经计算返回给用户的.用户可能会谎报请求以获得更低的价格完成传输任务.因此,RPM 的

设计需要考虑用户请求的真实性保证,具体的定义如下.

定义 2(真实性). 如果对于任意用户 $i \in \mathcal{N}$ 和其真实请求 q_i , 真实性要求对于任意其它的请求组合 $Q_i = \{q_{i1}, \dots, q_{im}\}$ 保证

$$u_i(\xi_i, d_i, \Delta t_i, C) \geq \sum_{j=1}^m u_i(\xi_{ij}, d_{ij}, \Delta t_{ij}, C) \quad (5)$$

从上面的不等式可知:对于任意一个传输任务,用户任何谎报请求的组合都不能比其真实的单个请求产生更高的收益.

2.4 强化学习

本文拟采用强化学习来设计定价机制,一方面可以在线学习用户的估价,另一方面面对不断变化的市场环境(例如带宽的剩余容量等)可以动态地调整定价策略.本小节将重点回顾多臂老虎机问题和 Q 学习算法.

2.4.1 多臂老虎机

多臂老虎机是一种经典的无状态强化学习问题.如果在时间 t 选择动作 a_t ,则会产生奖励 r_t .用 $v(a_t)$ 表示对动作 a_t 奖励的估计值,而 $v(a_t)$ 的真实值是探索学习的对象.最佳动作 a_t^* 是具有最高奖励的动作,即 $a_t^* = \arg \max_a v_t(a)$.理想最佳动作 a_t^* 的奖励 $v(a_t^*)$ 与实际动作 a_t 的奖励 r_t 之间的差值称为在时间 t 的悔恨.学习算法的优化目标是 minimized 所有 T 个时间的平均悔恨,即

$$\frac{1}{T} \sum_{t=1}^T (v(a_t^*) - r_t) \quad (6)$$

针对上述优化目标, ϵ -greedy 算法以 ϵ 的概率随机选择动作,否则选择经验平均奖励最高的动作.另一种经典的算法 Upper Confidence Bound(UCB)^[17] 则选择经验平均奖励和置信值上界之和最高的动作.在时间 t , UCB 算法的详细动作选择策略如下式所示:

$$a_t = \arg \max_a \left(r_{a,t-1} + l \sqrt{\frac{\ln t}{n_{a,t-1}}} \right) \quad (7)$$

其中 $r_{a,t-1}$ 表示到时间 $t-1$ 为止执行动作 a 的平均奖励, n_a 表示到时间 $t-1$ 为止选择动作 a 的次数, l 表示探索程度且取非负值.

2.4.2 Q 学习

Q 学习是由 Watkins 等人^[18]提出的一种无模型的强化学习方法,用以解决马尔可夫决策过程(Markov Decision Process, MDP)问题. MDP 问题包括 5 个关键参数:时间 t 时所处的状态 $s_t \in S$, 动作 $a_t \in A$, 状态转移概率 $P(s_{t+1} | s_t, a_t)$, 状态 s_t 下动

作 a_t 的奖励 r_t 和衡量未来奖励重要性的折扣因子 $0 \leq \gamma \leq 1$. Q 学习的目标是能够准确估计状态-动作组合收益的函数, 即 $Q: S \times A \rightarrow \mathbb{R}$. 在时间 t 、状态 s_t 下, 选择动作 a_t , 观察到奖励 r_t , 进入新状态 s_{t+1} . Q 函数将会采取以下方式更新:

$$Q(s_t, a_t) = (1 - \alpha_t) Q(s_t, a_t) + \alpha_t (r_t + \gamma \max_{a^*} Q(s_{t+1}, a^*)) \quad (8)$$

其中, α_t 表示在时间 t 的学习率, $\gamma \max_{a^*} Q(s_{t+1}, a^*)$ 表示在后续状态下的最高奖励乘以折扣因子 γ . 考虑到现实问题中状态空间的连续性或者大规模性, 可以将状态进行向量化表示, 并将 Q 参数化. 基于损失函数和随机梯度下降等优化算法来学习 Q.

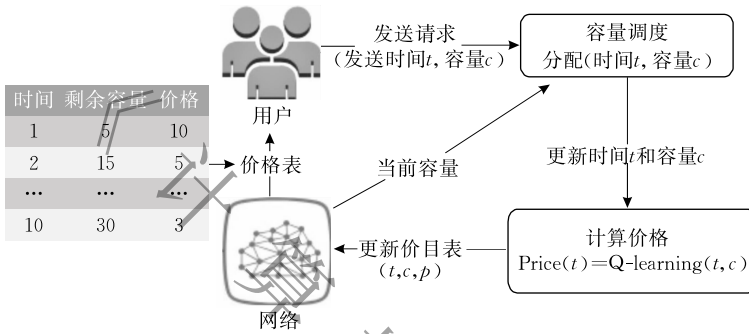


图 2 PLPM 的工作流程

3.1 设计细节

PLPM 将数据中心网络传输的在线定价建模为 MDP, 并将制定价格看作随机市场的决策过程. PLPM 定价机制中 MDP 建模的关键参数如下:

(1) 状态. 在时间 t , 服务提供商为任意的截止时长 $\Delta t \in \{1, \dots, \hat{t}\}$ 设定带宽的单位价格, 其中 $\hat{t}$ 是用户可选的最长截止时长. 用向量 $\mathbf{c} = (c_{\Delta t} | \forall \Delta t \in \{1, \dots, \hat{t}\})$ 表示时间 t 后续 $\hat{t}$ 个单位时间的剩余带宽容量. 为了确定时间 $t + \Delta t$ 的带宽价格, 我们考虑到: 截止时长 Δt 能刻画用户任务的紧迫程度, 而剩余的带宽容量 $c_{\Delta t} \in \{0, 1, \dots, c_e\}$ 能刻画网络状态, 其中 c_e 表示每个时间的最大带宽容量. 因此, 用在截止时长 Δt 的带宽容量 $c_{\Delta t}$ 作为 MDP 的状态.

(2) 动作. 将服务提供商的出价看作 MDP 中的动作. 用 $\bar{p} = \{\bar{p}_0, \bar{p}_1, \dots, \bar{p}_K\}$ 来表示候选价格的集合. 由于价格的范围是 $[\bar{p}, \hat{p}]$, 将第 k 个候选价格设置为 $\bar{p}_k = \bar{p}(1 + \nu)^k$, 其中 ν 是大于 0 的常数. 此外, 候选价格的数量 $K = \left\lfloor \log_{1+\nu} \frac{\hat{p}}{\bar{p}} \right\rfloor$. 因此, 动作空间的大小为 $K + 1$. 用 $p_{t, \Delta t}$ 表示在时间 t 、截止时长为 Δt 的价格.

(3) 奖励. PLPM 的目标是最大化服务提供商

3 PLPM 定价机制

本节介绍 PLPM 的设计细节, 工作流程如图 2 所示. 用户在某个时间进入市场时, 他们可以访问类似图中的价格表, 价格表中展示了每个时间可用的带宽容量以及该时间单位带宽的价格. 用户根据他们传输任务的截止时间需求和估价购买相应时间的带宽. 用户完成付款后, 服务提供商将为用户分配相应的带宽资源, 并将更新每个时间的剩余容量和价格. 为了适应未知和动态变化的市场环境, PLPM 主要采用 Q 学习算法更新价格表.

的累积收益. 因此, 在时间 $t + \Delta t$, 奖励是通过将价格设置为 $p_{t, \Delta t}$ 而获得的收入. 用 $d_{t, \Delta t}$ 表示当时售出的总容量, 可得奖励为 $r_{\Delta t} = p_{t, \Delta t} \times d_{t, \Delta t}$.

基于上述 3 个关键参数, 下面介绍 PLPM 基于 Q 学习的价格表更新策略. 在时间 t 的价格表包含 $\Delta t, c_{\Delta t}, p_{t, \Delta t}$ 三个部分. 将表中不同截止时长的单位带宽价格视为学习对象, 即总共有 $\hat{t}$ 个学习对象. 用 Q 值表 Q 来存储和更新 Q 函数值. 如算法 1 所示, 在时间 t , 服务提供商接收用户的传输请求, 为其分配带宽, 更新剩余容量, 并利用当前的价格表计算收益 (第 4~8 行). 本文为用户单次请求的传输量设置了上界 $\hat{d}$ (第 6 行), 防止传输量大或没有特定截止时间的任务 (例如后台传输 BT 任务) 占用所有资源, 同时为潜在的高估价任务预留带宽资源^[19]. 然后, 服务提供商根据当前收益来反向 (遍历截止时长 Δt 从 $\hat{t}$ 到 1) 更新 Q 值表 (第 9~12 行). 初始状态为超过最大的截止时长后 (即 $\hat{t} + 1$ 时间后) 后刷新的带宽容量 c_e (第 10 行). 此外, Q 值表以式 (8) 进行更新 (第 12 行). 最后采用 ϵ -greedy 策略更新 (在时间 $t + 1$ 的) 价格表, 即对于任意一个截止时长 Δt , 以 ϵ 的概率取随机价格, 否则取 Q 值表中最大的动作/价格 (第 13~18 行).

算法 1. PLPM 动态价格表机制.

1. 初始化

$$\Delta t \in [\hat{t}], s \in [c_e], p \in \bar{p}, Q(s, p), p_{1,\Delta t}, c_{\Delta t} = c_e$$

2. FOR $t = \{1, 2, \dots, T\}$ DO

3. $\forall \Delta t \in [\hat{t}], r_{\Delta t} = 0$

4. FOR 在时间 t 收到的请求 $q_t = (d, \Delta t)$ DO

5. //分配和更新网络带宽、更新收益

6. IF $d \leq c_{\Delta t}$ and $d \leq \hat{d}$ THEN

7. $c_{\Delta t} = c_{\Delta t} - d$

8. $r_{\Delta t} = r_{\Delta t} + p_{t,\Delta t} \times d$

9. //更新 Q 值表和价格表

10. $c_{t+1} = c_e$

11. FOR $\Delta t = \hat{t}$ to l DO

12. $Q(c_{\Delta t}, p_{t,\Delta t}) = (1 - \alpha_t) Q(c_{\Delta t}, p_{t,\Delta t}) + \alpha_t (r_{\Delta t} + \gamma \max_{p^*} Q(c_{\Delta t+1}, p^*))$

13. //用 ϵ -greedy 更新价格

14. 在 $[0, 1]$ 区间中选择随机数 ϵ' ;

15. IF $\epsilon' \leq \epsilon$ THEN

16. 将 $p_{t+1,\Delta t}$ 设为随机价格;

17. ELSE

18. $p_{t+1,\Delta t} = \arg \max_{p^*} Q(c_{\Delta t}, p^*)$

3.2 时间复杂度分析

假设在时间 t 收到 N_0 个传输请求. 鉴于每个用户基于价格表会(唯一地)选择使得价格最低的截止时长, 因此带宽分配的时间复杂度为 $O(N_0)$. 更新 Q 值表和设置价格的时间复杂度为 $O(\hat{t}_c K)$. 因此, PLPM 的时间复杂度为 $O(K \times \hat{t} \times c_e + n_t)$. 此外, 从 $K = \left\lceil \log_{1+\nu} \frac{\hat{p}}{\bar{p}} \right\rceil$ 个候选价格中选出的最优价格至少是

理论最优价格的 $\frac{1}{1+\nu}$. 因此, 参数 ν 可以用来调整定价算法收益和高效性之间的平衡.

3.3 PLPM 的不足

PLPM 允许用户选择进入带宽市场的时间和截止时间来满足网络传输需求. 然而, PLPM 只是简单地根据当前的网络状态来调整价格, 忽略了用户对于其传输任务的私有估价(即用户估价与传输类型、传输期限等因素的关系), 也就忽略了用户对于价格接受还是拒绝的反馈, 无法最大化服务提供商的累积收益. 此外, 由于传输时间和传输量是由用户选择的, 理性自私的用户将选择能保证截止时间内价格最低的传输方式, 而不管自己对于传输任务的真实估价. 这种策略性行为将对服务提供商造成巨大的收益损失.

4 RPM 定价机制

本节介绍 RPM 定价机制. 为解决 PLPM 的弊端, RPM 不仅考虑当前的网络状态, 而且还考虑用户提交的数据传输请求及用户对请求的估价. 图 3 展示了 RPM 的工作流程. 当用户在某个时间到达带宽网络市场时, 将提交数据传输请求并查询价格. RPM 根据不同传输类型(后台传输或者固定期限传输), 采用不同方法来计算价格并返回给用户. 用户将根据自己的估价决定接受还是拒绝价格. 如果用户接受, 容量调度模块会进行带宽分配. 首先, 服务提供商会赋予固定期限传输比后台传输更高的优先级. 之后, 将用户的传输量较均匀地分配在其截止时间之前, 以保证用户的传输要求. 分配完后, 相应的网络状态也会进行更新.

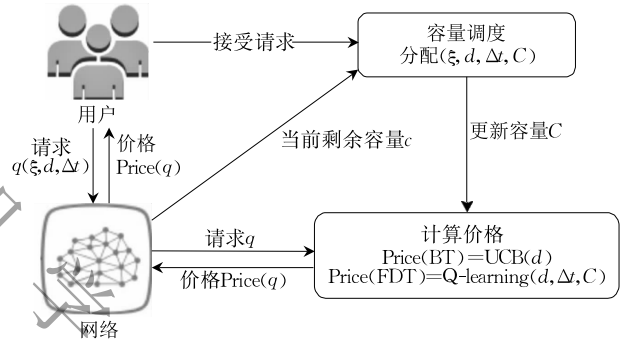


图 3 RPM 的工作流程

4.1 设计原理

鉴于难以直接探索动态变化的网络带宽市场, 同时也很难准确预测不同用户对于不同传输请求的估价, RPM 主要使用强化学习来为不同的请求进行定价. 首先, 后台传输任务没有明确的截止时间要求, RPM 使用无状态的 UCB 算法探索能最大化累积收益的价格. 对于固定期限传输任务, RPM 使用 Q 学习计算具有不同传输量、不同截止时间的请求在不同网络状态下的价格. 设定合适的定价函数, RPM 可以保证用户请求的真实性.

除了定价模块之外, RPM 的带宽容量调度模块可保证带宽有较高的利用率. 针对固定期限传输, 带宽分配的原则是在截止时间之前平衡可用容量, 从而降低网络拥塞的风险. 针对后台传输请求, 只有当网络带宽在下一个时间具有充足的剩余容量或该请求在特定时长里未获得带宽分配时, 服务提供商才为后台传输请求分配带宽. 上述的带宽分配策略为有更高的估价的用户保留了一些网络资源用于传

输,同时能避免后台传输任务无法完成。

最后考虑用户请求的真实性.用户可以谎报请求以增加收益.例如,把一个请求拆分成多个提交,请求一个比真实值更早或更晚的截止时间,或者将虚假的截止时间和虚假的传输量进行组合.因此,定价和分配规则应满足某些属性,以保证用户的真实性.为了更好地将强化学习方法与定价函数所需的属性结合在一起,RPM 首先确定一个基本的定价函数,其中包含需要用强化学习实现最高累积收益的函数参数,然后用 Q 学习来学习这些参数,而不是直接将价格设置为 Q 学习的动作.基于参数化 Q 学习的定价机制可以保证用户的真实性。

4.2 设计细节

后台传输请求的定价 后台传输任务通常有很宽裕的时间进行网络传输,甚至没有严格的截止时间.这意味着后台传输任务不要求在特定时间占有一定的带宽资源.因此,网络状态对用户估价没有影响.RPM 采用多臂老虎机中的 UCB 算法来探索最优价格,并使用与传输量成比例的定价函数.候选价格的集合为 $\bar{p} = \{\bar{p}_0, \bar{p}_1, \dots, \bar{p}_{K_B}\}$,设定策略与 PLPM 的保持一致.如算法 2 所示,服务提供商首先尝试每个候选价格,并获得相应奖励反馈(第 2~6 行).之后,根据 UCB 策略(即式(7))选择价格(第 7~12 行),该价格选择策略会倾向于选择经验收益更高或者被选择次数较少的候选价格。

算法 2. 面向后台传输(BT)任务、基于 UCB 算法的 RPM 定价机制

1. 初始化 $\forall 0 \leq k \leq K_B, n_k = 1, r_k = 0$
2. //探索每个候选价格
3. IF $1 \leq t \leq K_B + 1$ THEN
4. $p_t = \bar{p}_{t-1}$
5. d_t 量的带宽出售给当前时间 t 的用户
6. $r_{t-1} = p_t \times d_t$
7. //用 UCB 探索价格且更新最优价格
8. ELSE
9. $p_t = \bar{p}_k$, 其中 $k = \arg \max_k \left(r_k + l \sqrt{\frac{\ln t}{n_k}} \right)$
10. d_t 量的带宽出售给当前时间 t 的用户
11. $n_k = n_k + 1$
12. $r_k = \frac{p_t \times d_t + r_k \times (n_k - 1)}{n_k}$

固定期限传输请求的定价 在 RPM 中,对于固定期限传输任务的请求,服务提供商将在用户提交传输量和截止时间后计算价格.因此,用户有机会谎报他们的请求以获得更高的收益.面向固定期限传

输请求的价格计算包括两个部分:(1)找到可保证真实性定价函数的形式/性质;(2)使用强化学习来学习定价函数的最优参数设置,实现收益最大化的目标。

首先,我们给出如下的定理,以刻画保证用户请求真实性的定价函数。

定理 1. 满足以下性质的定价函数可以实现用户固定期限传输请求的真实性。

- (1) 价格非负. $\forall \Delta t \geq 0, d \geq 0, p(d, \Delta t) \geq 0$;
- (2) 价格与传输量呈单调增. $\forall d_1 > d_2 \geq 0, p(d_1) \geq p(d_2)$;
- (3) 定价函数 $p(d)$ 是凹函数;
- (4) 价格和截止时长成单调减. $\forall \Delta t_1 > \Delta t_2 \geq 0, p(\Delta t_1) \leq p(\Delta t_2)$;
- (5) 价格和截止时间的乘积呈单调增. $\forall \Delta t_1 > \Delta t_2 \geq 0, \Delta t_1 \times p(\Delta t_1) \geq \Delta t_2 \times p(\Delta t_2)$.

定理 1 中的前 3 个性质保证了用户真实地提交传输量,而后 2 个性质可以保证用户提交截止时间的真实性.定理 1 的详细分析证明将会在第 4.4 节给出。

在确定了可保证用户请求真实性的定价函数的形式后,直接使用 $q_i(\xi, d_i, \Delta t_i)$ 表示用户 i 的传输请求.与 PLPM 相一致,RPM 依然使用 Q 学习来进行动态定价.从第 3.1 节可知,计算复杂度随着状态的数量呈指数增长.因此,为了减少状态数量,使用当前时间至截止时间内平均剩余带宽作为 Q 学习的状态,即 q_i 的状态可表示为 $s_i = \frac{c_1 + c_2 + \dots + c_{\Delta t_i}}{\Delta t_i}$,

其中 Δt_i 不大于最长截止时长 $\hat{i}$.因此状态空间的大小为 $(c_e + 1)$.相比于 PLPM,RPM 主要的不同点在于对动作的定义.PLPM 直接将候选价格作为动作集,而 RPM 首先设定基本的定价函数,改变定价函数中的参数值可以改变定价函数,进而通过定价函数计算出价格,即 RPM 将候选的参数集当作动作集.首先,满足定理 1 的基本定价函数包括 $p(d, \Delta t) = \theta \frac{d}{\sqrt{\Delta t}}, p(d, \Delta t) = \theta_1 d + \theta_2 \frac{1}{\sqrt{\Delta t}}$ 等,用 $p(d, \Delta t; \theta)$ 来表示这一系列函数,其中 θ 表示储存所有参数的矩阵.根据价格的范围 $\bar{p} \leq \frac{p_i}{d_i} \leq \hat{p}$ 可得相应的参数的范围 $[\hat{\theta}, \bar{\theta}]$.将参数范围离散化为候选参数集(即动作集) $\{\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_{K_D}\}$,其中 $K_D = \left\lceil \log_{1+v'} \frac{\hat{\theta}}{\bar{\theta}} \right\rceil, \forall k \in 0, \dots, K_D, \hat{\theta}_k = \bar{\theta} (1 + v')^k$.因此,动作集的大小为

$K_D + 1$. 为了与状态的定义保持一致, RPM 中每个状态的奖励也需要被 Δt_i 平均. 因此, 奖励 $r(s_i, \theta_i) = \frac{r_i}{\Delta t_i} = \frac{p(d_i, \Delta t_i; \theta_i)}{\Delta t_i}$.

基于上述状态、动作和奖励的定义, 接下来介绍面向固定期限传输任务的定价机制. 如算法 3 所示, 当用户请求 q_i 到达时, 服务提供商首先确定状态 s_i (第 4 行), 然后选择具有最大 Q 值的动作 θ_i (第 11 行), 并使用定价函数 $p(d_i, \Delta t_i; \theta_i)$ 来计算价格 (第 12 行). 如果用户接受价格, 服务提供商将为其分配带宽容量 (第 14、15 行). 同时, 根据分配后的状态 s'_i 和奖励 r_i , 服务提供商将更新 Q 值表 (第 17 行). 如果用户拒绝该价格, 则此时的奖励为 0 并且网络带宽状态保持不变, 即下一状态等于此时状态, 更新 Q 值表 (第 19 行).

算法 3. 面向固定期限传输(FDT)任务、基于 Q 学习的 RPM 定价机制.

1. 初始化 $\forall s \in S, \theta \in \bar{\theta}, Q(s, \theta), p(d, \Delta t; \theta)$
2. FOR $t = \{1, 2, \dots, T\}$ DO
3. FOR 在时间 t 收到的用户请求 $q_i = (d_i, \Delta t_i)$ DO
4. 当前状态 $s_i = \frac{c_{t+1} + c_{t+2} + \dots + c_{t+\Delta t_i}}{\Delta t_i}$
5. IF $s'_i \geq 0$ THEN
6. //用 ϵ -greedy 选择价格
7. 在 $[0, 1]$ 区间中选择随机数 ϵ'
8. IF $\epsilon' \leq \epsilon$ THEN
9. 将 p_i 设为随机价格;
10. ELSE
11. $\theta_i = \arg \max_{\theta} Q(s_i, \theta)$
12. $p_i = p(d_i, \Delta t_i; \theta_i)$
13. //更新 Q 值表
14. IF 用户接受了价格 p_i THEN
15. 调用算法 4 进行带宽分配
16. 下一个状态 $s'_i = s_i - \frac{d_i}{\Delta t_i}$
17. $Q(s_i, \theta_i) = (1 - \alpha_t) Q(s_i, \theta_i) + \alpha_t \left(\frac{p_i}{\Delta t_i} + \gamma \max_{\theta^*} Q(s'_i, \theta^*) \right)$
18. ELSE
19. $Q(s_i, \theta_i) = (1 - \alpha_t) Q(s_i, \theta_i) + \alpha_t \gamma \max_{\theta^*} Q(s_i, \theta^*)$

容量分配 在带宽容量分配中, 固定期限传输任务的优先级高于后台传输任务. 首先, 将每个时间的总容量划分为 h 个单位, 每个单位容量为 $c' = \frac{c_e}{h}$. 按单位分配带宽, 并尝试使每个时间都有相近的容量, 从而使带宽余量在时间维度上相对平衡. 如算

法 4 所示, 对于每个时间 t , 服务提供商首先给固定期限传输任务分配带宽, 即在截止时间之前逐个填充容量直至满足传输需求为止 (第 6~11 行). 后台传输请求将添加到存储该类型请求的集合中, 并记录该请求上一次获得带宽分配的时间 (第 13 行). 在所有的固定期限传输请求处理完后, 空闲容量将分配给后台传输任务 (第 18~19 行). 为了避免由于网络繁忙而导致后台传输任务永远无法获得分配, 将未获得分配的次数上限设置为 $\hat{n}$. 如果某个后台传输请求 $\hat{n}$ 次尚未被分配任何的带宽, 将分配相应带宽给该请求 (第 20~23 行).

算法 4. 带宽分配机制.

1. 初始化后台传输集合 $B = \emptyset$
2. //在时间 t 给每个请求分配带宽
3. FOR $t \in \{1, 2, \dots, T\}$ DO
4. FOR 在时间 t 的每个用户请求 $q(\xi, d, \Delta t)$
5. $i = h$
6. //给固定期限传输请求分配带宽
7. IF $\xi = \text{FDT}$ THEN
8. WHILE $d > 0$ DO
9. $i = i - 1$
10. FOR $\Delta t_j = 1$ to Δt and $d > 0$ DO
11. 分配 $\max\{0, c_{\Delta t_j} - i \times c'\}$ 带宽
12. ELSE
13. $B = B \cup \{(q, cnt = 0)\}$
14. //分配未出售的带宽, 并考虑已经 $\hat{t}$ 时长未被分配带宽的后台传输请求
15. FOR $(q, cnt) \in B$ DO
16. IF $d = 0$
17. 从 B 中删除 (q, cnt) , 并将其它的 (q, cnt) 的 cnt 加 1
18. IF $c_1 > 0$ THEN
19. 从 c_1 带宽中分配至多 c' 的带宽给 q , 减去对应的 d , 然后重置 $cnt = 0$
20. FOR $(q, cnt) \in B$ 其中 $cnt \geq \hat{n}$ DO
21. $j = 2$
22. 从 c_j 带宽中分配至多 c' 给 q , 减去对应的 d , 重置 $cnt = 0$
23. $j = j + 1$

4.3 时间复杂度分析

假设服务提供商收到了 N_1 个后台传输请求和 N_2 个固定期限传输请求. 计算所有后台传输请求的价格需要 $O(K_B N_1)$ 时间, 而单个固定期限传输请求的定价需要为 $O(c_e K_D)$ 时间, 即总时间为 $O(c_e K_D N_2)$. 对于容量分配, 在最坏情况下, 后台传输和固定期限传输的分配分别需要 $O(N_1)$ 和 $O(h \hat{n} N_2)$. 因此, RPM

整体的时间复杂度为 $O(K_B N_1 + (c_r K_D + h \hat{N}_2))$.

4.4 真实性分析(定理 1 的证明)

用户的传输请求包含 3 个部分:传输类型、传输量和非后台传输类型请求中的截止时间. 本小节将从这三个方面及其组合来证明 RPM 的真实性.

传输类型 想要购买后台传输的用户没有动机将其提交为固定期限传输,因为它的价格会高很多. 具体来说,可以将后台传输看作具有最长截止时长的固定期限传输,再根据定理 1 中价格随截止时长单调递减的性质. 另一方面,由于后台传输无法提供截止时间保证,因此需要进行固定期限传输的用户也不可能谎报为后台传输.

传输量 为了确保传输请求中传输量的真实性,用户谎报传输量获得的收益不能大于真实报告获得的收益. 用 d_i 表示用户 i 真实所需的传输量,用 d_r 表示用户实际提交的传输量. 首先,用户不会提交比真实所需更少的总数据量,因为这样他们就无法完成传输任务;而如果用户报告的数据量超过真实所需,根据定义 1,他们将获得相同的估价,但根据定理 1,他们需要支付更高的价格,因此他们没有动机谎报传输量.

如果用户选择把一个任务分割成几个请求,多次提交的总传输数据量等于真实所需. 从定义 1 可知用户的估价 v_i 相同. 假设用户 i 提交的多个请求为 $Q_i = \{q_{i1}, q_{i2}, \dots, q_{im}\}$ 时,用 $D_i = \{d_{i1}, d_{i2}, \dots, d_{im}\}$ 表示多次请求的传输量并满足 $d_i = \sum_{j=1}^m d_{ij}$. 据定理 1 (定价函数 $p(d)$ 是凹函数)可知, $p(d_i) \leq \sum_{j=1}^m p(d_{ij})$. 用户提交真实传输量 d_i 的收益为 $u_i(d_i) = v_i - p(d_i) \geq v_i - \sum_{j=1}^m p(d_{ij}) = u_i(Q_i)$. 因此,用户没有动机提交多个请求以代替真实请求.

截止时长 由于后台传输没有明确的截止时间要求,因此仅讨论固定期限传输请求中截止时间的真实性. 用 Δt_i 表示真实的传输截止时长,用 Δt_r 表示用户报告的截止时长.

一种可能的情况是:用户向服务提供商提交更早的截止时间,即 $\Delta t_r < \Delta t_i$. 然而,根据定理 1 可知定价函数随着截止时间增大而递减,谎报更早的截止时间需要付更高的价格,即 $p_i(\Delta t_i) < p_i(\Delta t_r)$. 因此,用户将不会选择报告更早的截止时间. 另一种情况是:用户提交比真实截止时间更晚的时间,即 $\Delta t_r > \Delta t_i$,并做一次赌博希望在真实的截止时间前完成传

输. 直接分析这种情况非常困难. 例如, Jalaparti 等人^[8]讨论了用户的这种操作,但仅依据经验性的模拟结果得出:故意报告更晚的截止时间的潜在收益较低. 一般来说,如果用户提交比真实截止时间更晚的时间 Δt_r ,数据传输在真实的截止时间 Δt_i 之前完成任务的可能性,与开始时间至截至时间之间的空闲网络容量以及这段时间内新用户的传输任务密切相关. 用概率函数 $P(\Delta t_i, \Delta t_r, d_i, C)$ 表示这个可能性. 如果在用户的真实截止时间时传输仍未完成,假设在真实截止时间之前共传输了 d (可能是任何 $d < d_i$) 单位数据的概率为 $P_i(d, \cdot)$. 另外,将提交更晚的截止时间给用户带来的期望收益表示为 $\mathbb{E}(u_i(\Delta t_r))$. 可得保证传输截止时间真实性的引理.

引理 1. 给定定义 1 和定理 1,对任意用户 $i \in \mathcal{N}$ 的传输任务 q_i ,该用户没有动机谎报更晚的传输截止时间,因为无法从这种行为中获取更高的期望收益,即 $\forall \Delta t_r > \Delta t_i, \mathbb{E}(u_i(\Delta t_r)) \leq u_i(\Delta t_i)$.

证明. 用户谎报截止时间收益的期望值为:

$$\mathbb{E}(u_i(\Delta t_r)) = P(\Delta t_i, \Delta t_r, d_i, C) v_i(\xi_i, d_i, \Delta t_i) + (1 - P(\Delta t_i, \Delta t_r, d_i, C)) \int_0^{d_i} P_i(x, \cdot) v_i(\xi_i, x, \Delta t_i) dx - p_i(\Delta t_r) \quad (9)$$

由 $\forall d < d_i, v_i(\xi_i, d, \Delta t_i) = 0$ 可得:

$$\mathbb{E}(u_i(\Delta t_r)) = P(\Delta t_i, \Delta t_r, d_i, C) v_i(\xi_i, d_i, \Delta t_i) - p_i(\Delta t_r) \quad (10)$$

真实提交截止时间产生的收益为

$$u_i(\Delta t_i) = v_i(\xi_i, d_i, \Delta t_i) - p_i(\Delta t_i) \quad (11)$$

因此,要证 $\mathbb{E}(u_i(\Delta t_r)) \leq u_i(\Delta t_i)$ 需证:

$$1 - \frac{p_i(\Delta t_r)}{p_i(\Delta t_i)} \leq (1 - P(\Delta t_i, \Delta t_r, d_i, C)) \frac{v_i(\xi_i, d_i, \Delta t_i)}{p_i(\Delta t_i)} \quad (12)$$

另外,始终有 $u_i(\Delta t_i) = v_i(\xi_i, d_i, \Delta t_i) - p_i(\Delta t_i) \geq 0$ 且价格非负,可得:

$$\frac{v_i(\xi_i, d_i, \Delta t_i)}{p_i(\Delta t_i)} \geq 1 \quad (13)$$

因此,可得:

$$\frac{p_i(\Delta t_r)}{p_i(\Delta t_i)} \geq P(\Delta t_i, \Delta t_r, d_i, C) \quad (14)$$

假设用户不了解网络带宽容量状态以及如何安排其传输任务,数据将在 n 个时段里被传输,其中 $0 < n < \Delta t_r$. 由于我们的带宽分配算法是逐单位进行分配,可将概率简化为

$$P(\Delta t_i, \Delta t_r, d_i, C) = \max\left\{0, \frac{\Delta t_i - n}{\Delta t_r - n}\right\} \quad (15)$$

根据定理 1 的最后一个特性,即 $\Delta t_i \leq \Delta t_r, \Delta t_i \times p(\Delta t_i) \leq \Delta t_r \times p(\Delta t_r)$,可得: $\forall \Delta t_i > n > 0$,

$$\frac{p_i(\Delta t_r)}{p_i(\Delta t_i)} \geq \frac{\Delta t_i}{\Delta t_r} \geq \frac{\Delta t_i - n}{\Delta t_r - n} \geq P(\Delta t_i, \Delta t_r, d_i, C) \quad (16)$$

因此, $\forall \Delta t_r > \Delta t_i$, $\mathbb{E}(u_i(\Delta t_r)) \leq u_i(\Delta t_i)$ 始终成立, 用户没有动机提交更晚的截止时间. 证毕.

从上面的证明可以看出: 价格随截止时长的增加而减小以及截止时长和价格的乘积随截止时长的增大而增大, 保证了定价函数随截止时长递减的速度适当, 从而保证了用户请求中截止时间的真实性.

传输量与截止时间的组合 根据定义 1, 如果用户未能在规定的截止时间前完成传输任务, 则不会获得任何收益. 因此, 用户将寻求所有可能的方式在截止时间之前完成真实的传输量. 上面已分别讨论了用户请求中传输类型、传输数据量和截止时间的真实性. 接下来将分析用户同时谎报传输量和截止时间以获得更高收益的情况. 为了更好地说明这个问题, 举个简单的例子. $q = (d, \Delta t)$ 表示在时间 Δt 内传输 d 单位数据的请求. 用户 i 需要在 4 单位的时间内传输 10 单位的数据, 即 $q_i = (10, 4)$, 用户提交两个请求 $q_{i1} = (2, 3)$, $q_{i2} = (9, 5)$, 获得与 q_i 相同的估价, 并希望请求 q_{i2} 中 8 单位的数据可以在 4 单位时间内完成传输, 同时获得更低的总价格.

首先分析实际请求的传输总量等于真实所需传输量的情况. 用户 i 的真实请求为 $q_i = (d_i, \Delta t_i)$. 该用户将真实所需传输量 d_i 划分为任意(实际请求)传输量的集合 $D_i = \{d_{i1}, d_{i2}, \dots, d_{im}\}$, 并且满足 $d_i = \sum_{j=1}^m d_{ij}$. 提交这些传输量的请求集为 $Q_i = \{(d_{ij}, \Delta t_{ij}) \mid d_{ij} \in D_i\}$. 然后, 根据传输请求的截止时间与真实截止时间的关系, 将 Q_i 划分成 3 个不相交的请求子集: (1) 提交的截止时间在实际截止时间之前 $Q_{i, \text{before}} = \{(d_{ij}, \Delta t_{ij}) \mid d_{ij} \in D_i, \Delta t_{ij} < \Delta t_i\}$; (2) 提交的截止时间和真实的相同 $Q_{i, \text{true}} = \{(d_{ij}, \Delta t_{ij}) \mid d_{ij} \in D_i, \Delta t_{ij} = \Delta t_i\}$; (3) 提交的截止时间在真实的截止时间之后 $Q_{i, \text{after}} = \{(d_{ij}, \Delta t_{ij}) \mid d_{ij} \in D_i, \Delta t_{ij} > \Delta t_i\}$. 如果 $Q_{i, \text{after}}$ 中的请求在 Δt_i 时间之内没能完成, 则用户无收益. 因此, 仅需分析 $Q_{i, \text{before}}$ 和 $Q_{i, \text{true}}$ 给用户带来的收益:

$$\begin{aligned} u(Q_{i, \text{before}}) &= \sum_{(d_{ij}, \Delta t_{ij}) \in Q_{i, \text{before}}} u((d_{ij}, \Delta t_{ij})) \\ &\leq \sum_{(d_{ij}, \Delta t_{ij}) \in Q_{i, \text{before}}} u((d_{ij}, \Delta t_i)) \\ &\leq u\left(\sum_{d_{ij} \in Q_{i, \text{before}}} d_{ij}, \Delta t_i\right) \end{aligned} \quad (17)$$

$$\begin{aligned} u(Q_{i, \text{true}}) &= \sum_{(d_{ij}, \Delta t_i) \in Q_{i, \text{true}}} u((d_{ij}, \Delta t_i)) \\ &\leq u\left(\sum_{d_{ij} \in Q_{i, \text{true}}} d_{ij}, \Delta t_i\right) \end{aligned} \quad (18)$$

鉴于 $Q_{i, \text{after}}$ 中的每个请求 $(d_{ij}, \Delta t_{ij})$ 的期望收益都不会大于真实请求 $(d_{ij}, \Delta t_i)$ 的收益. 因此, 用户谎报请求的整体收益为

$$\begin{aligned} u(Q_i) &= u(Q_{i, \text{before}}) + u(Q_{i, \text{true}}) + u(Q_{i, \text{after}}) \\ &\leq u\left(\sum_{d_{ij} \in Q_{i, \text{before}}} d_{ij}, \Delta t_i\right) + u\left(\sum_{d_{ij} \in Q_{i, \text{true}}} d_{ij}, \Delta t_i\right) + \\ &\quad u\left(\sum_{d_{ij} \in Q_{i, \text{after}}} d_{ij}, \Delta t_i\right) \\ &\leq u\left(\sum_{d_{ij} \in Q_i} d_{ij}, \Delta t_i\right) \\ &= u(d_i, \Delta t_i) \end{aligned} \quad (19)$$

低于真实单个请求的收益. 因此, 用户不能通过谎报真实的传输总量来提高收益.

其次分析实际请求的传输总量大于真实所需传输量的情况, 即 $\sum_{j=1}^m d_{ij} > d_i$. 这种情况可以提高在真实的截止时间之前完成任务的可能性, 价格也可能更低. 与上面的分析相类似, $Q_{i, \text{before}}$ 和 $Q_{i, \text{true}}$ 不会出现谎报情况, 因此只需要考虑 $Q_{i, \text{after}}$, 即当 $d_r > d_i$ 和 $\Delta t_r > \Delta t_i$ 时, 需证明 $q_r = (d_r, \Delta t_r)$ 产生的收益小于请求 $q_i = (d_i, \Delta t_i)$. 下面给出的引理可得出如何保证用户不会提交晚于真实截止时间且超额传输量的请求.

引理 2. 给定定义 1 和定理 1, 对于任意用户 $i \in N$ 的传输任务 q_i , 该用户没有动机同时谎报较晚的截止时间和更高的传输量, 因为无法从这种行为获得更多收益, 即 $\forall d_r > d_i$ 和 $\Delta t_r > \Delta t_i$, $\mathbb{E}(u_i(d_r, \Delta t_r)) \leq u_i(d_i, \Delta t_i)$.

证明. 根据引理 1, 提交比真实情况更晚的截止时间和超额传输量, 用户的期望收益为

$$\mathbb{E}(u_i(d_r, \Delta t_r)) = P(\Delta t_i, \Delta t_r, d_i, d_r, C) v_i(\xi_i, d_i, \Delta t_i) - p_i(d_r, \Delta t_r) \quad (20)$$

用户真实的收益是:

$$u_i(d_i, \Delta t_i) = v_i(\xi_i, d_i, \Delta t_i) - p_i(d_i, \Delta t_i) \quad (21)$$

要证明 $\mathbb{E}(d_r, u_i(\Delta t_r)) \leq u_i(d_i, \Delta t_i)$ 即:

$$P(\Delta t_i, \Delta t_r, d_i, d_r, C) v_i(\xi_i, d_i, \Delta t_i) - p_i(d_r, \Delta t_r) \leq v_i(\xi_i, d_i, \Delta t_i) - p_i(d_i, \Delta t_i) \quad (22)$$

需要证明不等式 $\frac{v_i(\xi_i, d_i, \Delta t_i)}{p_i(d_i, \Delta t_i)} \geq 1$ 等价于证明

$$\frac{p_i(d_r, \Delta t_r)}{p_i(d_i, \Delta t_i)} \geq P(\Delta t_i, \Delta t_r, d_i, d_r, C) \quad (23)$$

根据定理 1 中价格和传输量成正比的性质, 可以得到 $p_i(d_i, \Delta t_r) \leq p_i(d_r, \Delta t_r)$, 所以只需证:

$$\frac{p_i(d_i, \Delta t_r)}{p_i(d_i, \Delta t_i)} = \frac{p_i(\Delta t_r)}{p_i(\Delta t_i)} \geq P(\Delta t_i, \Delta t_r, d_i, d_r, C) \quad (24)$$

可观察得出引理 1 和引理 2 证明之间的差别是概率大小. 在提出过多传输量的情况下, 截止时间内会分配更多的带宽, 这会提高在真实期限前完成真实传输量的可能性. 因此如果用户谎报成功, 即 d_i 数据在 n' 时间内完成传输, 且该时间在用户的真实期限内, 即 $0 < n' < n$. 可以将概率简化为

$$P(\Delta t_i, \Delta t_r, d_i, C) = \max \left\{ 0, \frac{\Delta t_i - n'}{\Delta t_r - n'} \right\}$$

此时, 仍然能保证 $\frac{\Delta t_i}{\Delta t_r} \geq \frac{\Delta t_i - n'}{\Delta t_r - n'}$. 于是, 证明和引理 1 的证明相同. 因此, 当定价函数满足定理 1 中的性质时, 可以保证 $\forall d_r > d_i$ 和 $\Delta t_r > \Delta t_i$, $\mathbb{E}(u_i(d_r, \Delta t_r)) \leq u_i(d_i, \Delta t_i)$ 一直成立. 因此, 用户没有动机提交超出真实截止时间且更高传输量的请求. 证毕.

根据上面的分析, 具有定理 1 中 5 个性质的定价函数具有鲁棒性: 理性自私的用户无法通过谎报请求获得比真实请求更高的收益. 因此, RPM 保证了网络带宽传输市场的真实性.

5 实验结果

本节将展示 PLPM 和 RPM 的实验结果, 并主要从服务提供商的累积收益和网络带宽利用率两方面与其他定价机制进行比较.

5.1 实验设置

网络结构 实验采用谷歌跨数据中心的 B4 广域网拓扑结构. 该网络具有 12 个数据中心和 19 条链路^[20], 每条链路的容量基于引力模型进行估算, 大约是 80 Gbps 的整数倍^[6]. 因此在实验中, 将每条链路的单位容量设置为 80.

传输请求 根据微软对广域网上用户的调查^[8], 实验假设每个时间用户的传输请求中包含 80% 的固定期限传输 (FDT) 请求和 20% 的后台传输 (BT) 请求. 针对固定期限传输请求, 在 $\{1, 2, \dots, \hat{t}\}$ 中随机选择作为用户提交的截止时长. 因为传输量和传输时间有关, 将两者比值, 即传输速率 $d/\Delta t$, 设置为服从期望为 5 的指数分布. 针对后台传输请求, 将传输数据量设置为满足期望为 60 的指数分布.

评估指标 实验主要考虑服务提供商的累积收益以及网络带宽的利用率. 为了验证用户请求的真实性, 还计算了用户的预期收益. 在网络带宽市场中, 可行的价格区间设置为 $[1, \hat{p}]$. 针对固定期限传输请求, 考虑两种类型的用户估价函数: (1) 估价仅

与传输量有关, 与截止时间无关. 采用线性的估价函数 $v(d, \Delta t) = \beta_1 d$, 其中 $\beta_1 = 0.6\hat{p}$; (2) 价值不仅和传输量成正相关, 还和截止时长呈负相关. 采用估价函数 $v(d, \Delta t) = (\beta_2 - \beta_3 \Delta t) d$, 其中 $\beta_2 = 0.9\hat{p}$, $\beta_3 = 0.6\hat{p}/\hat{d}$. 针对后台传输请求, 将估价函数设置为 $v(d) = \beta_4 d$, 其中 $\beta_4 = 0.4\hat{p}$. 鉴于实际中不存在确定型的用户估价, 实验中用户的估价从高斯分布 $N(v, \sigma^2)$ 上进行随机采样以验证定价机制的鲁棒性, 其中 $\sigma = 0.05\hat{p}$.

定价机制 为了表现 PLPM 和 RPM 的优势, 实验引入一些基准性的定价机制进行对比, 并介绍所有定价机制中详细的参数设置.

(1) 固定价格 (简称 FIX). 这是目前服务提供商主流的定价机制 (例如表 1). 实验将链路中每单位带宽的价格设置为 $\frac{1+\hat{p}}{2}$. 此外, 通过使用带宽分配算法 4 来提高网络带宽的利用率.

(2) 离线最优 (简称 OPT). 理想化的 OPT 将价格设置为用户的真实估价.

(3) PLPM. Q 学习算法的参数设置为 $\alpha = 0.1$, $\epsilon = 0.9$, $\gamma = 0.8$. 每个时间请求数据量的上界为 $d = 30$. 考虑到时间复杂度和整体收益表现, 参数 $\nu = 0.15$.

(4) RPM. 定价函数设为 $p(d, \Delta t; \theta) = \theta \frac{d}{\sqrt{\Delta t}}$.

鉴于每单位数据传输量的价格区间是 $[1, \hat{p}]$, 可获得定价函数的参数范围 $\theta \in [\sqrt{\hat{t}}, \hat{p}]$. 同样, 将容量分成 $h = 8$ 个单位, 每单位有 $c' = 10$ 个容量进行分配. RPM 中 Q 学习算法的其它参数与 PLPM 相同. UCB 算法的探索率设定为 $l = 20$.

5.2 性能表现

首先评估当时间从 1 增加到 2000 时不同定价机制的收益表现. 图 4(a) 显示了累积收益, 而图 4(b) 显示不同定价机制的累积收益与最优累积收益之间的比值. 从图 4(a) 中, 可以观察到: 所有定价机制的累积收益与时间成线性关系, 且 RPM 是除 OPT 之外增长速度最快的. 从图 4(b) 中, 可以观察到: 在经过 2000 单位的时长后, RPM 的总收益大约为 OPT 的 80%, 而 FIX 大约只有 OPT 的 20%. 此外, 还可以观察到: RPM 在初始阶段存在一些波动, 这是 UCB 和 Q 学习算法的探索阶段所引起的, 但强化学习的算法可以保证快速地学习/收敛到较优的策略. 例如, 在 500 个单位时间时, RPM 的累积收益

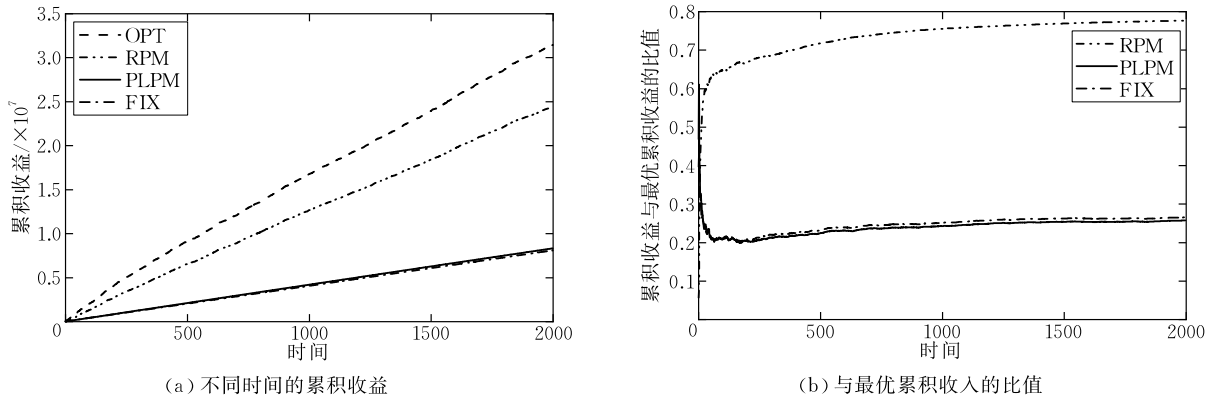


图 4 不同定价机制在不同时间的收益表现

已经达到 OPT 的 70%左右;并在 1500 单位时长后收敛.值得注意的是,初期的探索对于定价策略的优化是必不可少,因为如果没有初期的探索,也就无法进行后续精准的利用,无法实现长期累积收益的优化.

其次测试当价格上界从 20 增加到 200、增量为 20 时不同定价机制的表现.在两种不同的用户估价函数下,经过 1000 单位的时长后,对 50 次的运行结果取平均,结果如图 5 所示.可以看到:无论用户的估价是否与截止时间有关,RPM 的累积收益总是高于 FIX 和 PLPM.此外,RPM 的累积收益大约是 OPT 的 2/3.这表明本文所提出的定价机制以较低收益损失确保了交易市场的真实性.同时,上述实验结果也表明了 RPM 允许选择相对少量的候选价格来提高市场效率.鉴于实验中的 FIX 针对用户的估价分布设定了近优的固定价格并配备了基于优先级的带宽分配策略,它的总收益要稍优于 PLPM.这也侧面反映了服务提供商学习用户估价对于最大化其累积收益的重要性.

再次,假定用户达到服从泊松分布 (Poisson distribution),并测试不同用户到达率(泊松分布的

期望和方差,即每个时间到达带宽市场的平均用户数量)设定下不同定价机制的累积收益.当用户到达率从 5 开始,以 5 为单位增加到 50 时,不同定价机制的累积收益如图 6 所示.鉴于带宽市场随着用户到达率的增加而趋于饱和, FIX 和 PLPM 的累积收益几乎保持不变.相比之下, RPM 的累积收益与用户的到达率近似成线性增长.具体地,如图 7 所示,当用户到达率为 5 时, FIX 和 PLPM 的带宽利用率已经分别大于 60%和 80%,而 RPM 的带宽利用率已经饱和;当用户到达率达到 10 时,所有定价机制的带宽利用率都已饱和.此外,当用户到达率从 5 增加到 10 时, FIX 和 PLPM 由于带宽利用率的增加而产生的累积收益增量,与 OPT 和 RPM 累积收益的增量相比相对较小,无法在图中显著体现.因此,图 6 中 FIX 和 PLPM 的累积收益曲线近乎直线.从上述实验观察结果和分析中可以得出: RPM 能在带宽资源受限(即用户到达率高、带宽趋近饱和)的情况下较好地挑选最有价值的用户以优化服务提供商的累积收益.另外, RPM 和 OPT 累积收益的比较结果与在不同价格上界情况下基本一致.

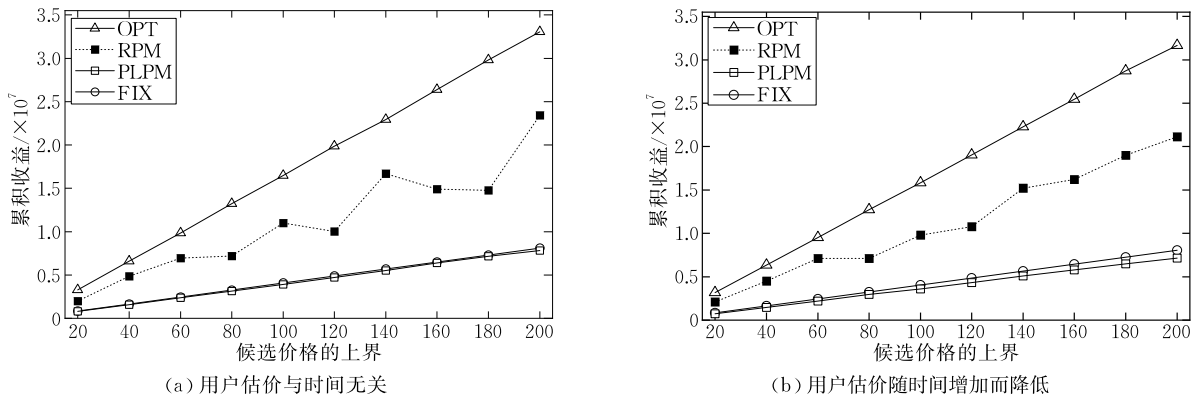


图 5 在不同价格上界设置下的累积收益(经过 1000 单位的时长后)

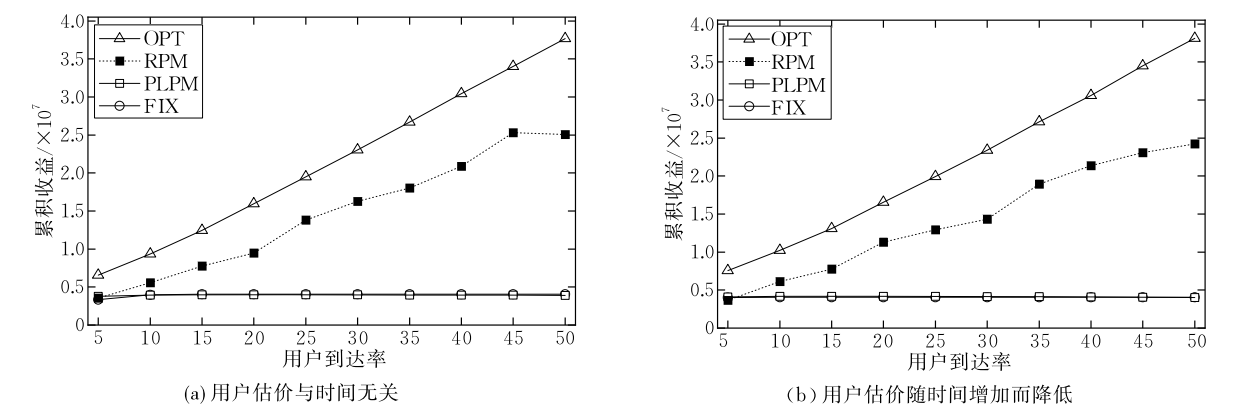


图 6 在不同用户到达率设定下的累积收益(经过 1000 单位的时长后)

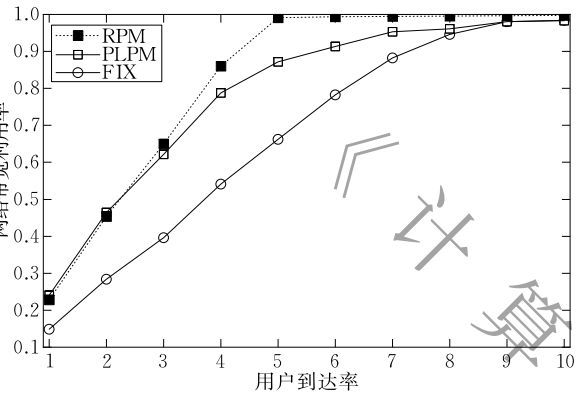


图 7 不同定价机制的带宽利用率

实验还比较了不同定价机制下的带宽利用率,结果如图 7 所示.当用户请求量太大而无法满足不同用户请求时,带宽利用率将达到 100%,因此选择到达率较小的设定.具体地,将到达速率的范围设置为 1 到 10、增量为 1 以覆盖空闲和繁忙的网络带宽状态.鉴于使用不同估价函数和参数得到的结果相似,因此选择随截止时间递减的用户估价函数.从图 7 中,可以观察到:在所有情况下,RPM 的性能皆为最佳,因为分配规则的设计可确保收益最大化的同时让网络容量有较高的利用率.此外,PLPM 在网络资源分配的方面表现胜过 FIX,因为 PLPM 在不同的网络状态下具有动态价格.具体来说,在用户到达率不高的情况下,FIX 网络带宽利用率较低的原因在于,固定的价格损失了一些估价较低的用户.相比之下,PLPM 和 RPM 均会采用降低价格的方式使得过剩的网络带宽资源得到充分利用,同时提高累积收益.相比于 PLPM,RPM 进一步学习了传输类型、传输数据量和截止时间等因素对用户估价的影响,可以更准确地预测用户估价.因此,在用户到达率不高的情况下,RPM 的网络带宽利用率优于 FIX 和 PLPM,并较快地使网络带宽达到饱和.

最后验证在 RPM 机制下用户请求的真实性.从第 4.4 节(引理 1 和引理 2)可知:如果定价函数满足真实性,用户虚假请求的价格将比其真实请求的价格更高,用户收益降低.具体到本实验,对于提交比真实截止时间更晚的请求或同时提交超额传输量的请求,验证用户的实际收益降低.由于两种不同真实性的实验得到了相似的结果,因此仅报告当提交更晚截止日期时用户收益的期望值.实验测试截止时长从 2 到 5 的请求,观察用户提交晚于真实截止时间的时长与收益期望值的关系,结果如图 8 所示.如果值为零,则表示用户提交了真实的请求.实验结果将 1000 次测试结果进行平均.由于并非每次用户都会被分配到带宽,因此当以 1000 次进行平均时,观察到,真实请求的预期收益接近于零,但仍然为正;对比之下,对于每个具有较晚截止时间的请求,用户都会产生较大的负收益.实验结果表明:RPM 对延长截止时间的策略性行为具有鲁棒性,用户将没有动机谎报他们的请求.

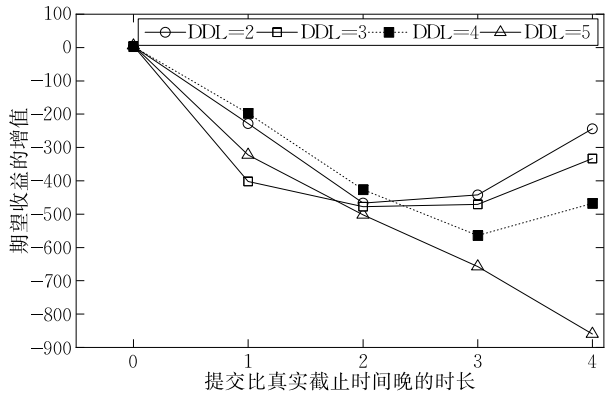


图 8 提交更晚截止时间产生的预期效益增值

总结来说,PLPM 和 RPM 根据网络带宽剩余容量进行动态的价格调整,提高了带宽利用率.此外,RPM 考虑了用户细粒度的传输请求并对用户估

价进行了在线学习,显著提高了服务提供商的累积收益.最后,RPM 对用户潜在的谎报行为具有鲁棒性,保证了真实性.

6 相关工作

本节回顾国内外已有相关工作.

6.1 数据中心间网络带宽市场

面向数据中心间广域网带宽资源的研究已经引起了工业界和学术界的高度关注.大型云服务提供商已采用软件定义网络(Software-Defined Networking, SDN)^[21]以提高带宽利用率,例如谷歌的 B4 及其变种^[3,20]、微软的 SWAN^[6]等.特别地,Kumar 等人^[22]设计了灵活、层次化的带宽分配方案来提高带宽利用率.然而,这些工作仅为不同的传输任务分配了优先级,并使用流量工程(Traffic Engineering)^[23]技术来分配资源,但没有传输截止时间的保证.有一些工作^[24-26]讨论了数据中心间的传输保证,但是它们不够灵活,仅在带宽容量较小时提供保证.Kandula 等人^[27]提出了 Tempus 框架,该框架能最大化在截止时间前完成传输任务的比例.Zhang 等人^[28]则考虑如何最大化传输任务的完成量.李文信^[29]在其博士论文中指出:充分利用在空闲时间的流量可以削减传输成本,同时灵活地放置网络流的端点可以减少跨数据中心 Coflow 的传输完成时间.上述研究工作都使用流量工程的工具进行了带宽分配以提升网络利用率,而本文则侧重基于实时的网络带宽状态对用户的传输请求进行定价,并为接受价格的用户分配带宽资源,即以动态的价格杠杆引导带宽资源分配.此外,本文中在线定价机制的设计主要考虑如何保证网络经济学的相关性质,包括累积收益最大化和用户请求的真实性.

智能数据定价(Smart Data Pricing,SDP)^[30]相关的研究已经有数十年之久了.Sen 等人^[31]对此进行了全面的文献调研.智能数据定价是指使用定价机制来管理网络资源、改善网络状况的方法,主要目标是减少网络拥塞、提升用户体验、提高服务提供商的经济效益等.例如,傅晓明等人^[32]根据网络资源的供需关系,设计了以优化用户满意度和系统效率为目标的定价策略.姜永等人^[33]将移动终端用户与运营商之间的数据定价和分配建模成多主多从的 Stackelberg 博弈模型,并提出了最大化多方效用的机制.请参阅更多相关的文献^[34-38]了解更详细的信息.已经存在一些工作将智能数据定价的相关技

术用于为数据中心间的广域网带宽设计定价机制.Zhao 等人^[39]考虑了面向数据中心间任务调度和资源配置的动态定价.Zheng 等人^[5]提出了针对不同网络拓扑结构、具有多项式时间复杂度的定价机制,并能实现近似收益最大化.但是,他们在问题模型中简化地考虑了网络链路无容量限制的理想情况.Yang 等人^[40]考虑了服务提供商如何从用户传输请求的集合中挑选出部分请求进行带宽分配以最大化收益.相比之下,本文所提出的定价机制考虑了网络链路的容量限制,且服务每个时间点到达市场的全集用户.Jalaparti 等人^[8]通过将动态定价和流量工程相结合,以最大化社会福利为目标,提供可以保证截止时间的网络传输服务.相似的工作还有^[41].与上述工作不同的地方在于,本文优化的目标是最大化服务提供商的累积收益.

真实性(也称激励相容)是机制设计中一个重要的概念^[42].如果一种机制可以激励参与者采取与自己内心意愿相一致的动作(例如定价机制中用户私有的估价),则该机制保证了真实性.已有大量研究考虑如何为在线拍卖机制提供真实性保证.例如,Babaioff 等人和 Devanur 等人的研究^[43-44]发现:基于 MAB 的定价机制无法同时保证真实性和累积收益最大化.也有一些工作^[45-49]在网络资源分配场景下利用在线拍卖机制来获得不同资源真实的优先级.魏蛟龙和张弛^[50]考虑了简单的单时刻带宽分配问题,并提出采用统一价格拍卖保证诚实性.在该研究工作,价格与网络带宽状态无关,并与用户的传输请求无关.苑迎等人^[51]指出了统一价格拍卖在激励效果和公平性等方面存在的问题,提出了基于非完全信息博弈的云资源分配模型,并利用隐马尔可夫预测服务提供商的出价.郑臻哲等人^[52]针对预留带宽的分配问题,设计了保证用户真实性并最大化社会福利的拍卖机制.然而,拍卖机制无法适用于要求低时延的数据中心间带宽资源的实时分配.还有一些工作^[3,6,8,16,28]已经关注了用户对服务的多层次、细粒度的需求.然而,这些工作没有考虑用户行为的真实性问题.Jalaparti 等人^[8]经验性地说明:当在网络所有链接上都采用最低价格时,用户谎报请求很难提升其收益.本文设计了面向数据中心间网络带宽的在线定价机制,并在理论上保证了用户请求的真实性.

6.2 在线定价机制设计

动态的市场交易过程中,卖方可以实时地获取关于买方的信息/反馈和市场的状态,在线定价机制

由此产生。当卖方对于市场的需求知之甚少或者不确定性很高时, 在线定价则尤为必要^[53]。在理论计算机科学(Theoretical Computer Science)领域中, 在线定价的相关工作主要依赖于知道需求分布或者买方的估价分布以探索商品的最优价格。工作^[54-56]主要使用贝叶斯学习来探索在有限时间内的最优定价策略。Cohen 等人^[53]测试了定价机制在线性需求曲线下和其他任意需求分布下的性能差距。他们的后续工作^[57]摆脱了对于分布假设的依赖, 提出了基于采样方法实现收益最大化。Cesa-Bianchi 等人^[58]针对一般性的 MAB 问题优化了累积遗憾。还有一些工作(例如文献^[59-60])关注了商品供应有限而需求高度不确定的情况, 并重点考虑如何优化在不同时间的供应量。上述方法主要适用于同一类商品且买家对其的估价类似, 而无法适用于我们所考虑的用户可自定义商品且用户估价不同的场景。还有一些工作面向用户估价动态变化的情况设计了在线定价机制。例如, Mirrokni 等人^[61-62]研究了具有不确定未来估价分布的重复拍卖机制, Mao 等人^[64]针对估价随时间衰减的商品设计了在线出价拍卖机制。然而, 这些工作未考虑对估价产生影响的因素, 即在本文 MDP 问题模型下的状态, 包括带宽容量、传输类型、传输量、截止时长等。

6.3 与国内外研究工作的比较

首先, 与基于流量工程的带宽分配工作相比, 本文的主要不同点在于, 通过设计动态的定价机制实现价格杠杆引导网络资源分配。其次, 与智能数据定价、数据中心网络带宽定价、保证真实性的网络资源定价等方面现有工作的主要不同点在于, 本文考虑在线低时延、网络链路存在带宽容量上界的市场模型, 并以最大化服务提供商的累积收益同时保证用户真实性为目标, 而非最大化社会福利。最后, 与理论上现有的在线定价机制相比, 本文的主要不同点在于用户可自定义商品(即传输请求)且用户估价不同。此外, 本文利用强化学习的相关工具建模了与用户估价相关的具体因素, 学习了实时的网络带宽容量和用户具体的传输请求与用户估价之间的关系。

7 结论与未来工作

本文研究了数据中心网络带宽的在线定价机制, 考虑了如何满足用户的个性化传输请求、如何保证用户请求的真实性以及如何最大化服务提供商的累积收益。首先提出了基于强化学习的 PLPM, 从供

求关系的角度对数据传输请求进行定价, 从而向用户提供了在不同时间、不同带宽容量状态下的单位带宽价目表。进一步提出了 RPM。RPM 是一种根据用户传输请求内容确定价格的机制, 并针对特定的传输类型设计了特定的定价方案和分配规则。具体来说, 对于后台传输, RPM 使用了无状态的强化学习算法来学习用户的估价, 并且在网络空闲时进行数据传输; 对于固定期限传输请求, RPM 使用参数化的 Q 学习算法计算具有不同传输量、不同截止时间的请求在不同网络状态下的价格, 在定价函数的形式上保证了用户请求的真实性, 并在截止时间之前均匀地分配带宽以满足传输任务的需求。最后, 实验验证了所提出的在线定价机制在累积收益、带宽利用率和用户真实性方面性能的优越性。

下一步的研究可以考虑用户决策是否接受服务提供商价格的时延, 并研究决策时延对带宽分配和动态定价的影响。潜在的解决思路包括: 服务提供商预先设置用户决策时延的上界, 并依据该上界设计带宽预分配算法和带宽预定价机制。

参 考 文 献

- [1] Luo L, Foerster K, Schmid S, et al. DaRTree: Deadline-aware multicast transfers in reconfigurable wide-area networks//Proceedings of the International Symposium on Quality of Service(IWQoS 2019). Phoenix, USA, 2019, 28: 1-10
- [2] Luo L, Foerster K, Schmid S, et al. Deadline-aware multicast transfers in software-defined optical wide-area networks. IEEE Journal on Selected Areas in Communications, 2020, 38(7): 1584-1599
- [3] Hong C, Mandal S, Al-Fares M, et al. B4 and after: Managing hierarchy, partitioning, and asymmetry for availability and scale in Google's software-defined WAN//Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication (SIGCOMM). Budapest, Hungary, 2018: 74-87
- [4] Adhikari V K, Guo Y, Hao F, et al. Unreeling netflix: Understanding and improving multi-CDN movie delivery//Proceedings of the 2012 IEEE Conference on Computer Communications (INFOCOM 2012). Orlando, USA, 2012: 1620-1628
- [5] Zheng Z, Srikant R, Chen G. Pricing for revenue maximization in inter-datacenter networks//Proceedings of the 2018 IEEE Conference on Computer Communications (INFOCOM 2018). Honolulu, USA, 2018: 28-36
- [6] Hong C, Kandula S, Mahajan R, et al. Achieving high utilization with software-driven WAN//Proceedings of the ACM SIGCOMM 2013 Conference(SIGCOMM). Hong Kong, China, 2013: 15-26

- [7] Chen Y, Jain S, Adhikari V K, et al. A first look at inter-datacenter traffic characteristics via Yahoo! datasets//Proceedings of the 2011 IEEE Conference on Computer Communications (INFOCOM 2011). Shanghai, China, 2011: 1620-1628
- [8] Jalaparti V, Bliznets I, Kandula S, et al. Dynamic pricing and traffic engineering for timely inter-datacenter transfers//Proceedings of the ACM SIGCOMM 2016 Conference. Florianopolis, Brazil, 2016: 73-86
- [9] Narahari Y, Raju C V L, Ravikumar K, et al. Dynamic pricing models for electronic business. *Sadhana*, 2005, 30(2): 231-256
- [10] Rana R, Oliveira F S. Real-time dynamic pricing in a non-stationary environment using model-free reinforcement learning. *Omega*, 2014, 47: 116-126
- [11] Wu T. Automated pricing agents in the on-demand economy: UCB/EECS-2016-57 [Ph. D. dissertation]. EECS Department, University of California, Berkeley, USA, 2016
- [12] Ruelens F, Claessens B J, Vandaal S, et al. Residential demand response of thermostatically controlled loads using batch reinforcement learning. *IEEE Transactions on Smart Grid*, 2017, 8(5): 2149-2159
- [13] Tang P. Reinforcement mechanism design//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne, Australia, 2017: 5146-5150
- [14] Mao W, Zheng Z, Wu F, et al. Online pricing for revenue maximization with unknown time discounting valuations//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 440-446
- [15] Ahani G, Wiatr P, Yuan D. Routing and scheduling of network flows with deadlines and discrete capacity allocation. *Networks*, 2020, 76(1): 54-74
- [16] Wilson C, Ballani H, Karagiannis T, et al. Better never than late: Meeting deadlines in datacenter networks//Proceedings of the ACM SIGCOMM 2011 Conference. Toronto, Canada, 2011: 50-61
- [17] Auer P, Cesa-Bianchi N, Fischer P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 2002, 47: 235-256
- [18] Watkins C J, Dayan P. Technical note: Q-learning. *Machine Learning*, 1992, 8: 279-292
- [19] Bogle J, Bhatia N, Ghobadi M, et al. TEAVAR: Striking the right utilization-availability balance in WAN traffic engineering //Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM 2019). Beijing, China, 2019: 29-43
- [20] Jain S, Kumar A, Mandal S, et al. B4: Experience with a globally deployed software defined wan//Proceedings of the ACM SIGCOMM 2013 Conference (SIGCOMM). Hong Kong, China, 2013: 3-14
- [21] Laoutaris N, Sirivianos M, Yang X, et al. Inter-datacenter bulk transfers with NetStitcher//Proceedings of the ACM SIGCOMM 2011 Conference. Toronto, Canada, 2011: 74-85
- [22] Kumar A, Jain S, Naik U, et al. BwE: Flexible, hierarchical bandwidth allocation for WAN distributed computing//Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication (SIGCOMM). London, UK, 2015: 1-14
- [23] Reda W, Bogdanov K, Milolidakis A, et al. Path persistence in the cloud: A study of the effects of inter-region traffic engineering in a large cloud provider's network. *ACM SIGCOMM Computer Communication Review*, 2020, 50(2): 11-23
- [24] Guo C, Lu G, Wang H J, et al. SecondNet: A data center network virtualization architecture with bandwidth guarantees //Proceedings of the 2010 ACM Conference on Emerging Networking Experiments and Technology (CoNEXT). Philadelphia, USA, 2010: 1-12
- [25] Ballani H, Costa P, Karagiannis T, et al. Towards predictable datacenter networks//Proceedings of the ACM SIGCOMM 2011 Conference. Toronto, Canada, 2011: 242-253
- [26] Xie D, Ding N, Hu Y C, et al. The only constant is change: Incorporating time-varying network reservations in data centers //Proceedings of the ACM SIGCOMM 2012 Conference (SIGCOMM). Helsinki, Finland, 2012: 199-210
- [27] Kandula S, Menache I, Schwartz R, et al. Calendaring for wide area networks//Proceedings of the ACM SIGCOMM 2014 Conference (SIGCOMM), Chicago, USA, 2014: 515-526
- [28] Zhang H, Chen K, Bai W, et al. Guaranteeing deadlines for inter-data center transfers. *IEEE/ACM Transactions on Networking*, 2017, 25(1): 579-595
- [29] Li Wen-Xin. Research on Techniques of Bandwidth Allocation and Flow Scheduling in Large-Scale Data Center [Ph. D. dissertation]. Dalian University of Technology, Dalian, 2018 (in Chinese)
(李文信. 大规模数据中心带宽分配与流量调度技术研究[博士论文]. 大连理工大学, 大连, 2018)
- [30] Clark D D. Internet cost allocation and pricing//Internet Economics. Hawaii, USA: MIT Press, 1997: 215-252
- [31] Sen S, Joe-Wong C, Ha S, et al. A survey of smart data pricing: Past proposals, current plans, and future trends. *ACM Computing Surveys*, 2013, 46(2): 15:1-15:37
- [32] Fu Xiao-Ming, Zhang Yao-Xue, Ma Hong-Jun, et al. A market based approach to allocate bandwidth for computer networks. *Acta Electronica Sinica*, 1999, 27(9): 127-132(in Chinese)
(傅晓明, 张尧学, 马洪军等. 一种基于市场模型的网络带宽分配方法. *电子学报*, 1999, 27(9): 127-132)
- [33] Jiang Yong, Chen Shan-Zhi, Hu Bo. Stackelberg games-based distributed algorithm of pricing and resource allocation in heterogeneous wireless networks. *Journal on Communications*, 2013, 34(1): 61-68(in Chinese)

- (姜永, 陈山枝, 胡博. 异构无线网络中基于 Stackelberg 博弈的分布式定价和资源分配算法. 通信学报, 2013, 34(1): 61-68)
- [34] Sen S, Joe-Wong C, Ha S, et al. Smart data pricing: Using economics to manage network congestion. *Communications of the ACM*, 2015, 58(12): 86-93
- [35] Basar T, Srikant R. Revenue-maximizing pricing and capacity expansion in a many-users regime//*Proceedings of the 2002 IEEE Conference on Computer Communications (INFOCOM 2002)*. New York, USA, 2002: 294-301
- [36] Marbach P, Berry R A. Downlink resource allocation and pricing for wireless networks//*Proceedings of the 2002 IEEE Conference on Computer Communications (INFOCOM 2002)*. New York, USA, 2002: 1470-1479
- [37] Hande P, Chiang M, Calderbank A R, et al. Pricing under constraints in access networks; Revenue maximization and congestion management//*Proceedings of the 2010 IEEE Conference on Computer Communications (INFOCOM 2010)*. San Diego, USA, 2010: 938-946
- [38] Agarwal A, Dahleh M A, Sarkar T. A marketplace for data: An algorithmic solution//*Proceedings of the 2019 ACM Conference on Economics and Computation (EC 2019)*. Phoenix, USA, 2019: 701-726
- [39] Zhao J, Li H, Wu C, et al. Dynamic pricing and profit maximization for the cloud with geo-distributed data centers//*Proceedings of the 2014 IEEE Conference on Computer Communications (INFOCOM 2014)*. Toronto, Canada, 2014: 118-126
- [40] Yang Z, Cui Y, Wang X, et al. Less is more; Service profit maximization in geo-distributed clouds. *IEEE Transactions on Cloud Computing*, 2020. DOI: 10.1109/TCC.2020.3024616
- [41] Chawla S, Miller J B, Teng Y. Pricing for online resource allocation: Intervals and paths//*Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2019)*. San Diego, USA, 2019: 1962-1981
- [42] Hurwicz L. On the concept and possibility of informational decentralization. *The American Economic Review*, 1969, 59(2): 513-524
- [43] Babaioff M, Sharma Y, Slivkins A. Characterizing truthful multi-armed bandit mechanisms; Extended abstract//*Proceedings of the 10th ACM Conference on Electronic Commerce*. Stanford, USA, 2009: 79-88
- [44] Devanur N R, Kakade S M. The price of truthfulness for pay-per click auctions//*Proceedings of the 10th ACM Conference on Electronic Commerce*. Stanford, USA, 2009: 99-106
- [45] Hajiaghayi M T, Kleinberg R D, Mahdian M, et al. Online auctions with re-usable goods//*Proceedings of the 6th ACM Conference on Electronic Commerce*. Vancouver, Canada, 2005: 165-174
- [46] Godfrey B, Schapira M, Zohar A, et al. Incentive compatibility and dynamics of congestion control//*Proceedings of the 2010 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*. New York, USA, 2010: 95-106
- [47] Shnayder V, Parkes D C, Kawadia V, et al. Truthful prioritization for dynamic bandwidth sharing//*Proceedings of the Fifteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*. Philadelphia, USA, 2014: 235-244
- [48] Awerbuch B, Azar Y, Meyerson A. Reducing truth-telling online mechanisms to online optimization//*Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC)*. San Diego, USA, 2003: 503-510
- [49] Feigenbaum J, Papadimitriou C H, Sami R, et al. A BGP-based mechanism for lowest-cost routing. *Distributed Computing*, 2005, 18(1): 61-72
- [50] Wei Jiao-Long, Zhang Chi. Research on auction-based bandwidth allocation for computer networks. *Chinese Journal of Electronics*, 2003, 31(6): 891-894(in Chinese)
(魏蛟龙, 张驰. 基于拍卖的网络带宽分配方法的研究. 电子学报, 2003, 31(6): 891-894)
- [51] Yuan Ying, Wang Cui-Rong, Wang Cong, et al. An uncompleted information game based resources allocation model for cloud computing. *Journal of Computer Research and Development*, 2016, 53(6): 1342-1351(in Chinese)
(苑迎, 王翠荣, 王聪等. 基于非完全信息博弈的云资源分配模型. 计算机研究与发展, 2016, 53(6): 1342-1351)
- [52] Zheng Zhen-Zhe, Wu Fan, Chen Gui-Hai. On designing multi-dimensional strategy-proof auctions for distributed cloud bandwidth reservation. *Chinese Journal of Computers*, 2019, 42(4): 701-720(in Chinese)
(郑臻哲, 吴帆, 陈贵海. 多维度防策略性云带宽预留拍卖机制设计. 计算机学报, 2019, 42(4): 701-720)
- [53] Cohen M C, Perakis G, Pindyck R S. Pricing with limited knowledge of demand//*Proceedings of the 2016 ACM Conference on Economics and Computation*. Maastricht, The Netherlands, 2016: 657
- [54] Araman V F, Caldentey R. Dynamic pricing for nonperishable products with demand learning. *Operations Research*, 2009, 57(5): 1169-1188
- [55] Besbes O, Zeevi A J. Dynamic pricing without knowing the demand function; Risk bounds and near-optimal algorithms. *Operations Research*, 2009, 57(6): 1407-1420
- [56] Agrawal S, Wang Z, Ye Y. A dynamic near-optimal algorithm for online linear programming. *Operations Research*, 2014, 62(4): 876-890
- [57] Cohen M C, Lobel R, Perakis G. Dynamic pricing through data sampling. *Production and Operations Management*, 2018, 27(6): 1074-1088
- [58] Cesa-Bianchi N, Gentile C, Mansour Y. Regret minimization for reserve prices in second-price auctions. *IEEE Transactions on Information Theory*, 2015, 61(1): 549-564
- [59] Badanidiyuru A, Kleinberg R, Slivkins A. Bandits with knapsacks. *Journal of the ACM*, 2018, 65: 13:1-13:55
- [60] Badanidiyuru A, Langford J, Slivkins A. Resourceful contextual bandits//*Proceedings of the 27th Conference on Learning Theory (COLT)*. Barcelona, Spain, 2014: 1109-1134

[61] Mirrokni V S, Leme R P, Tang P, et al. Non-clairvoyant dynamic mechanism design//Proceedings of the 2018 ACM Conference on Economics and Computation. Ithaca, USA, 2018: 169

[62] Mirrokni V S, Leme R P, Tang P, et al. Dynamic auctions with bank accounts//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI). New York, USA, 2016: 387-393



NIU Chao-Yue, Ph.D. His research interests include data sharing/trading, mobile computing, machine learning, and privacy and security.

CHEN Pei-Yu, M. S. candidate. His research interests include data pricing and network economics.

ZHANG Jia-Yi, M. S. Her research interests include network economics and reinforcement learning.

WU Fan, Ph.D., professor. His research interests include network economics, wireless networking, mobile computing, and privacy and security.

CHEN Gui-Hai, Ph.D., professor. His research interests include distributed computing and computer networks.

Background

In recent few years, the worldwide demands for cloud-based services are growing sharply, and the bandwidth on the Wide Area Network (WAN) that connects different data-centers is becoming more and more valuable. For example, web search engines require high-throughput transfers to update and propagate new indices; video delivery services require low-latency transfers to improve user experience; big data applications need to transfer large-scale datasets from one datacenter to another, thus performing the tasks of data integration, data mining, machine learning, and deep learning. Realizing such huge demands of data transfer, many domestic and overseas cloud platforms (e. g., Amazon Web Services (AWS), Google Cloud, Microsoft Azure, Alibaba Cloud, Huawei Cloud, and Tencent Cloud) have built up globally distributed datacenters as well as WAN to connect them for data transfer services.

In this work, we consider how to dynamically price inter-datacenter bandwidth, from the perspective of a service provider. We first identify the main drawbacks in the current pricing strategy with fixed prices and the bandwidth allocation strategy with traffic engineering: (1) the data transfers with strict deadlines cannot be supported; (2) the cumulative revenue of the service provider cannot be maximized; and (3) the network bandwidth cannot be adequately utilized. We thus turn to design online pricing mechanisms and overcome the following three major challenges: (1) how to build a suitable market model to capture the complex supply-demand relationships between the service provider and customers? (2) how to online exploit and explore the private valuations of customers for maximizing the service provider’s

cumulative revenue? and (3) how to guarantee the truthfulness of customers?

We propose two reinforcement learning-based pricing mechanisms, called Price List-based online Pricing Mechanism (PLPM) and Request-Based Online Pricing Mechanism (RPM). First, PLPM provides a price list for customers in a real-time way. The price list contains time slots as well as remaining capacities and prices in correspondence. By contrast, RPM allows the customers to launch customized requests for data transfer. RPM further online learns the highly differentiated valuations of customers and ensures their truthfulness with parameter-based Q-learning. To improve network utilization, we propose a priority-based strategy for bandwidth allocation. We finally evaluate our design under different valuation functions of customers. Experimental results reveal the remarkable advantage of our design over the pricing strategy with fixed prices from higher cumulative revenue and network utilization. Specifically, the cumulative revenue of RPM converges to roughly 80% of the offline optimum.

This work was supported in part by the Science and Technology Innovation 2030—“New Generation Artificial Intelligence” Major Project (No. 2018AAA0100900), in part by the China NSF Grant Nos. 62025204 and 61972254, and in part by the Alibaba Group through Alibaba Innovation Research (AIR) Program. The opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies or the government.