

# 基于深度确定性策略梯度的 MPTCP 数据调度算法

韩佳龙 滕建霖

西南民族大学 计算机与人工智能学院, 四川成都, 610225;

**摘要:** 多路径 TCP (MPTCP) 是由互联网工程任务组设计的多路径传输协议, 允许设备使用多个路径并行传输数据。数据调度算法负责按照调度规则将数据分配给不同路径传输。研究证明, 数据调度算法对 MPTCP 的传输性能有很大影响。现有调度算法在动态、异构网络环境下存在一定的局限性, 没有解决乱序数据包数量问题, 且大多算法只能提高 MPTCP 的一种传输性能。针对上述问题, 通过将深度确定性策略梯度引入数据调度问题, 使算法适应复杂的网络环境, 并通过多目标奖励函数提升多个传输性能。实验证明, 改进后的算法在乱序数据包数量、吞吐量和往返时延等方面皆优于对比调度算法。

**关键词:** MPTCP; 数据调度算法; 深度确定性策略梯度

**DOI:** 10.69979/3041-0673.25.07.016

## 引言

随着网络的发展, 多路径传输协议因其可以建立多条路径并行传输数据而受到人们的关注, 多路径传输控制协议 (Multipath TCP, MPTCP) 作为 TCP 的多路径扩展协议成为了研究热点。数据调度机制是多路径传输协议独有的机制, 该机制决定如何将传输数据分配给不同的路径, 在提高网络资源利用率的同时, 提高 MPTCP 的传输性能。

当前调度算法分为启发式调度算法和基于深度强化学习调度算法两类, 前者以 MinRTT<sup>[1]</sup>、BLEST<sup>[2]</sup>为代表, 通过固定的调度规则来分配数据, 调度策略无法根据网络环境的变化进行相应的调整, 导致该类算法无法适应网络的异构性、动态性, 在传输过程中出现大量乱序 (OFO) 数据包, 进而出现队头 (Ho1) 阻塞和接收缓冲区限制, 影响 MPTCP 的传输性能。

后者将深度强化方法应用于数据调度问题, 催生如 Peekaboo<sup>[5]</sup>以及 DEAR<sup>[6]</sup>等数据调度算法。该类算法可根据网络环境制定自适应调度策略, 弥补了启发式算法无法适应异构、动态网络环境的问题, 但现有算法所使用的深度强化学习方法并不适合 MPTCP 数据传输环境, 且只能提升一种传输性能。

针对上述问题, 本文提出基于深度确定性策略梯度的数据调度算法 DDPGBS, 将深度确定性策略梯度应用于数据调度问题, 使算法可以适应复杂的网络场景, 并通过设计多目标奖励函数, 减少了乱序数据包数量, 提高了吞吐量, 降低了往返时延。

## 1 基于深度确定策略梯度调度算法设计

如图 1 所示, 算法分为两部分:

(1) 数据调度部分: 调度器将观察网络环境, 并将吞吐量、往返时延、拥塞窗口大小等作为评估路径传输能力的状态特征输入 Actor 神经网络, 网络输出对应调度动作, 即分配给每条路径的数据量。调度器收集当前状态 $s_t$ 、调度动作 $a_t$ 、奖励 $r_t$ 以及下一个状态 $S_{t+1}$ 等信息, 将其组成四元组存入经验缓冲区供算法训练使用。

(2) 训练部分: 当缓冲区中的调度经验达到一定数量后, 智能体随机地从缓冲区中抽取经验数据训练 Actor 和 Critic 网络。Critic 网络对 Actor 网络输出的策略进行评估, 根据 Q 网络的损失函数更新其网络参数。Actor 网络依照 Critic 网络对其的评估, 根据梯度上升的方法更新其网络参数, 通过交替更新 Critic 网络和 Actor 网络相应参数来实现最优调度策略的学习。为了减少样本相关性, 避免 Q 值估计震荡, 训练部分采取了经验回放机制以及双网络机制。使得在训练部分算法共有四个神经网络: 主、目标 Actor 网络和主、目标 Critic 网络。四个网络的功能分别为:

主 Actor 网络: 将当前网络环境的状态信息作为输入, 输出对应的数据调度动作。

目标 Actor 网络: 将网络环境的下一个状态信息作为输入, 输出与下一状态对应的数据调度动作。

主 Critic 网络: 对当前状态下选择的调度动作进行策略评估, 计算其 Q 函数估值。

目标 Critic 网络: 对下一状态对应的调度动作进行策略评估, 计算其 Q 函数估值, 并计算当前状态动作的 Q 目标值。

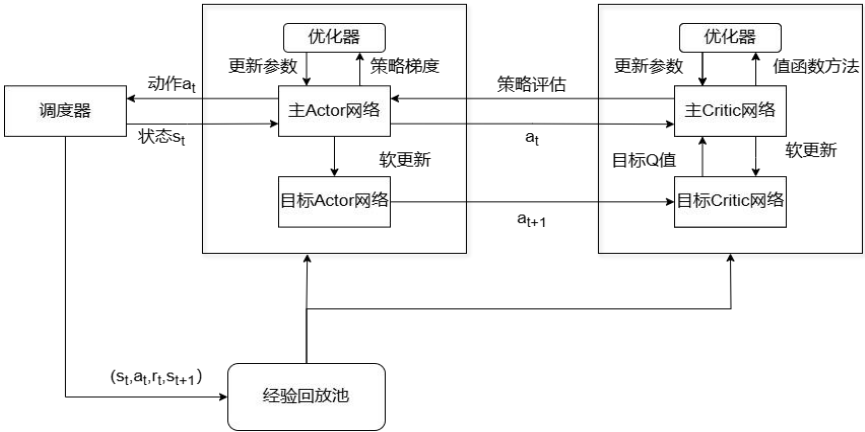


图 1 调度算法框架图

2 奖励函数的设计

奖励函数是调度器评估调度策略采取动作好坏的重要依据。在强化学习任务中，智能体的目标是最大化累计奖励值，因此我们设计一个多目标奖励函数，考虑了多种 Qos 特性，使算法在传输过程中减少乱序数据包数量，提高吞吐量，降低往返时延，将 MPTCP 的传输性能最大化，奖励函数如下：

$$R(s_t, a_t) = \alpha V^{\text{th}} - \beta V^{\text{RTT}} - \gamma V^{\text{QF0}}$$

其中， $0 \leq \alpha, \beta, \gamma \leq 1$  表示各个网络性能的权重，

用于平衡函数中每个传输性能的大小。

3 实验结果与分析

3.1 实验环境搭建

实验在 Linux 平台通过 Mininet 模拟器进行，采用 Ubuntu16.04 操作系统，Pythorch 深度学习框架，实验设计的网络拓扑结构如图 2 所示。其中客户端有两个网络接口，服务器有一个网络接口，客户端通过两条部分不相交的路径连接到服务器。

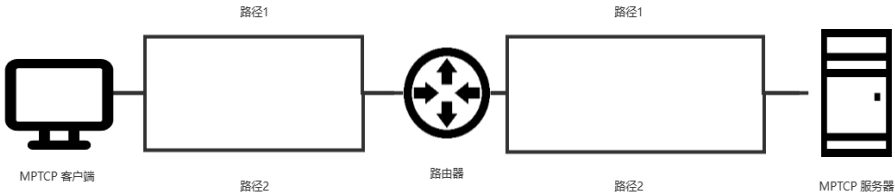


图 2 网络拓扑图

模拟环境中网络特征的具体参数如表 3-1 所示，路径特征的选择是在现实网络环境中对 LTE 和 Wifi 网络大规模测量得到的，实验通过客户端从服务器下载一个 32MB 的文件来测试算法的传输性能，每个实验重复 20 次以其平均值作为最终结果。

表 1 测试网络环境中路径参数设置范围

| 路径 | 带宽<br>(Bandwidth) | 往返时延<br>(RTT) | 丢包率<br>(Loss) |
|----|-------------------|---------------|---------------|
| 1  | 33                | 22            | 0.1%          |
| 2  | 20                | 50            | 0.1%          |

3.2 评价标准

本文以乱序数据包数量、吞吐量、往返时延大小为标准来评估算法的性能，对比调度算法为 MinRTT、BLEST 和 Peekaboo，涵盖了启发式数据调度算法和基于深度强化学习数据调度算法。

3.3 实验结果

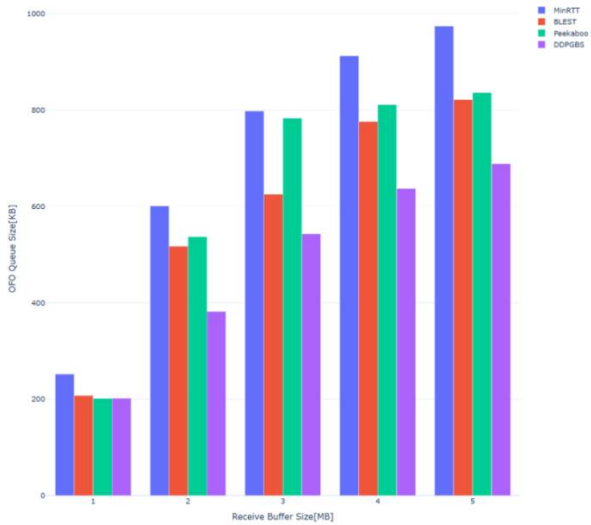


图 3 乱序数据数量对比

实验将乱序数据包数量作为评判标准,如图3所示,在接收缓冲区容量不同的情况下,乱序数据包数量对比算法减少了约17%~25%。

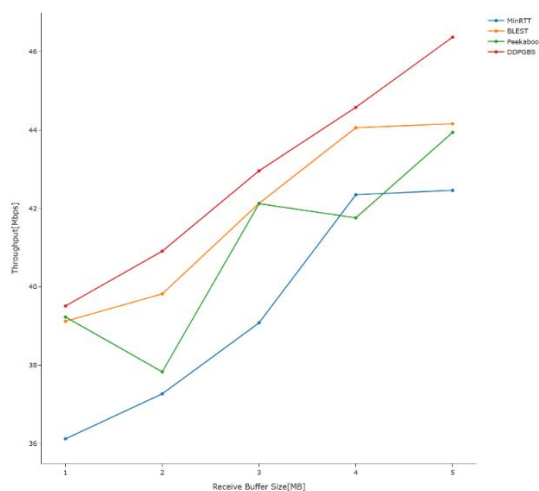


图4 吞吐量对比

实验将吞吐量大小作为评判标准,如图4所示,在接收缓冲区容量不同的情况下,算法所提供的吞吐量高于对比算法4%~11%。

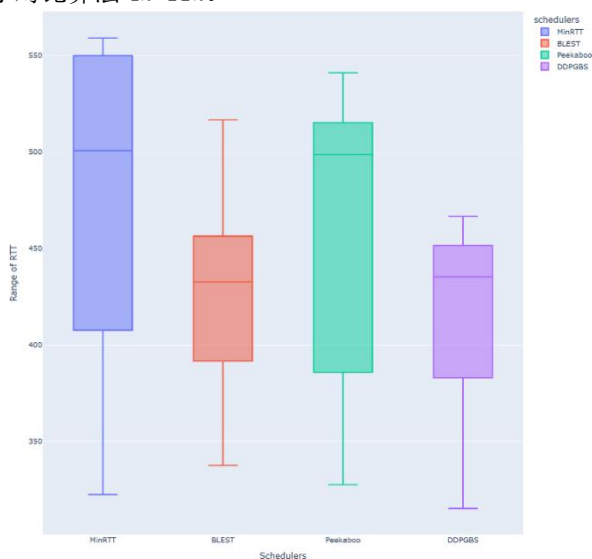


图5 RTT 范围对比

实验将往返时延作为评判标准,如图5所示,算法往返时延的范围在316.7ms~466.3ms,最大值明显低于其他调度算法,且绝大部分往返时延小于450ms,视频流媒体应用程序对RTT的要求应小于450ms,如果大于450ms,将会严重降低用户的应用体验,算法基本满足视频流媒体应用对往返时延的要求。

## 4 结语

本文提出了基于深度确定性策略梯度的数据调度算法,通过将深度确定性策略梯度引入MPTCP数据调度场景,使算法可以更好的适应异构、动态网络环境,设计了多目标奖励函数,在传输过程中,减少了乱序数据包数量、提高了吞吐量、降低了往返时延。实验结果表明,在模拟网络环境中,本文提出的DDPGBS数据算法在乱序数据包数量、吞吐量和往返时延等方面皆优于对比调度算法。

## 参考文献

- [1]C. Paasch, S. Ferlin, O. Alay, and O. Bonaventure, "Experimental evaluation of multipath tcp schedulers," in Proceedings of the 2014 ACM SIGCOMM workshop on Capacity sharing workshop, 2014, pp. 27 - 32.
- [2]S. Ferlin, O. Alay, O. Mehani, and R. Borli, "Blest: Blocking estimation-based mptcp scheduler for heterogeneous networks," in 2016 IFIP networking conference (IFIP networking) and workshops. IEEE,
- [3]H. Wu, O. Alay, A. Brunstrom, S. Ferlin, and G. Caso, "Peekaboo: Learning-based multipath scheduling for dynamic heterogeneous environments," IEEE J. Sel. Areas Commun., vol. 38, no. 10, pp. 2295 - 2310, Oct. 2021.
- [4]Kumar, P. S., Saxena, P., Alay, Ö. (2024). DEAR: DRL Empowered Actor-Critic Scheduler for Multipath QUIC Under 5G/B5G Hybrid Networks. In: Barolli, L. (eds) Advanced Information Networking and Applications. AINA 2024. Lecture Notes on Data Engineering and Communications Technologies, vol 199. Springer, Cham.

## 作者简介:

韩佳龙(1995-),男,锡伯族,吉林敦化人,硕士研究生,研究方向为MPTCP数据调度。

滕建霖(1999-),男,汉族,四川宜宾人,硕士研究生,本文通信作者,研究方向为MPTCP拥塞控制。