

基于多目标优化的网络拥塞控制算法的研究分析^{*}

奚文杰^{1,2} 汪小威² 李陶深^{1,2*}

1. 南宁学院信息工程学院, 南宁 530200
2. 广西大学计算机与电子信息学院, 南宁, 530004

摘要 在高动态和异构的网络环境下, 传统的拥塞控制算法在动态适应性和多目标优化能力方面逐渐显现出局限性。基于深度强化学习的拥塞控制方法通过动态调整策略和优化资源分配, 为提高吞吐量、降低延迟和改善公平性提供了全新解决方案。本文阐述了传统拥塞控制算法的发展现状, 以及基于强化学习的拥塞控制研究进展, 分析公平性优化在拥塞控制中的作用, 对基于深度强化学习的拥塞控制方法的未来研究工作给出一些研究思路。

关键字 拥塞控制, 多目标优化, 强化学习, 车-路协同, 车-车协同, 车-物协同

Research and Analysis of Network Congestion Control Algorithm Based on Multi Objective Optimization

Xi Wenjie^{1,2} Wang Xiaowei² Li Taoshen^{1,2**}

1. School of Information Engineering, Nanning University, Nanning 530200, China
2. School of Computer, Electronics and Information, Guangxi University, Nanning 530004, China
lts@unn.edu.cn

Abstract—In highly dynamic and heterogeneous network environments, traditional congestion control algorithms gradually show limitations in terms of dynamic adaptability and multi-objective optimization capabilities. The congestion control method based on deep reinforcement learning provides a new solution for improving throughput, reducing latency, and enhancing fairness by dynamically adjusting policies and optimizing resource allocation. This paper elaborates on the current development status of traditional congestion control algorithms, the research progress of congestion control based on reinforcement learning, and analyzes the role of fairness optimization in congestion control. It also provides some research ideas for future research on congestion control methods based on deep reinforcement learning.

Keywords—Congestion control, multi-objective optimization, reinforcement learning, ,

1 引言

拥塞控制是调节流量源数据传输速率以有效利用网络容量的核心网络任务, 随着实时视频、虚拟现实、物联网等互联网服务的出现, 拥塞控制受到了广泛关注。

近年来, 随着互联网流量的迅猛增长和网络应用的不断演进, 传统的拥塞控制 (Congestion control, CC) 算法面临着越来越多的挑战。在高动态和异构的网络环境下, 现有基于规则的拥塞控制方法 (如 CUBIC 算法^[1]和 BBR 算法^[2]) 难以适应复杂多变的网络条件, 尤其是在需要同时优化多目标性能 (如吞吐量、时延和丢包率) 以及确保公平性的场景下。为了应对这些问题, 研究者们开始探索基于深度强化学习 (Deep Reinforcement Learning, DRL) 的拥塞控制方法, 其凭借自适应能力和动态优化潜力展现出广阔的应用前景。

然而, 目前基于 DRL 的拥塞控制方法仍存在一些亟待解决的问题。首先, 现有的方法在优化网络性能时往往倾向于单一目标, 如最大化吞吐量或最小化时延, 忽略了多目标优化中的权衡关系。这种单一目标的优化策略可能导致流量间的不公平性, 进而影响用户体验和网络资源的合理利用^[3-4]。其次, 多目标优化问题 (如同时最小化时延和能耗) 在实际网络环境中通常是相互冲突的, 无法同时达到最优。此外, 现有研究大多基于静态的训练环境, 缺乏在动态网络条件下的实时适应能力, 进一步限制了算法的实际部署和应用, 如 TCP-RL 算法^[5]、DeepCC 算法^[6]。

在拥塞控制算法的设计中, 资源优化是一个核心问题。传统的强化学习算法由于需要频繁的决策更新, 往往带来较高的计算和资源开销, 这对实时性和实际部署构成了重大挑战。为此, 双重控制机制 (Dual Control Mechanism, DCM) 作为近年来备受关注的一种技术路径, 它通过引入低频决策与高频控制的分离, 显著降低了系统的资源开销。具体而言, 该机制将高

* 基金资助: 本文得到广西重点科技攻关项目 (桂科 AD20297125)

** 通信作者: 李陶深 lts@unn.edu.cn

频事件（如拥塞窗口的调整）交由传统拥塞控制算法处理，而将低频决策（如传输速率的长期优化）交给强化学习代理完成，实现了传统规则与现代机器学习技术的有机结合。

可以说，深度强化学习驱动的拥塞控制方法的研究作为当前网络优化的重要研究方向，已经引起了国内外学者的广泛关注与探索。基于深度强化学习的拥塞控制方法能够动态地调整策略和优化资源分配，为提高吞吐量、降低延迟和改善公平性提供了全新解决方案。本文将介绍传统拥塞控制算法的发展现状，基于深度强化学习的拥塞控制研究进展，分析公平性优化在拥塞控制中的作用，并介绍我们在该领域的未来研究的一些研究思路。

2 传统拥塞控制算法的研究现状

在网络环境中，为了向应用程序提供高效且可靠的传输服务，需要实施端到端传输控制。但网络拥塞的发生会导致吞吐量下降、数据分组丢失、延迟增加以及网络资源的低效利用，网络拥塞是指当数据流量超过网络承载能力时，网络无法提供预期服务的状态。当主机向网络发送超出其可处理能力的数据包时，链路和路由器缓冲区可能会发生拥塞，进而导致吞吐量急剧下降、数据包丢失、传输延迟增加，以及带宽和计算资源的严重浪费。因此，拥塞控制算法被广泛用于优化网络传输性能，提升可靠性。拥塞控制算法的核心目标是在最大化网络吞吐量的同时，尽可能避免过载导致的数据丢失和长时间排队。此外，一个有效的拥塞控制机制不仅应能适应动态变化的网络环境，还需确保在多个竞争用户/节点之间实现端到端的公平资源分配，以维持整体网络的稳定性和公平性。

传统的拥塞控制算法分为两大类：端到端拥塞控制^[7-8]和网络辅助拥塞控制^[9]。端到端控制方法仅依赖于发送方和接收方的协作，因此无需网络中显式信号的支持。而网络辅助控制方法需要网络设备的支持，例如路由器提供的拥塞信息，这在复杂的网络场景中实现公平性和响应性方面尤为重要。

端到端拥塞控制的核心挑战在于如何利用网络隐式信号（如数据包丢失、传输时延、带宽估计等）准确识别或预测网络拥塞。当前，端到端拥塞控制算法主要分为三类：基于丢包的拥塞控制、基于时延的拥塞控制、基于带宽估计的拥塞控制、混合拥塞控制以及基于机器学习的拥塞控制。

基于丢包的信号方法通过检测在特定时间内发送方未收到相应确认（ACK）的情况调整发送速率，这通常表明发生了数据包丢失。当网络设备的缓冲区超载时会产生丢包，因此此类方法通过利用链路带宽来实现高吞吐量。例如，文献[9]提出了用于分组交换网络拥塞避免的随机早期检测网关，该网关通过计算平

均队列大小来检测初期拥塞，通过丢弃到达网关的数据包或在数据包报头中设置一个比特来通知连接拥塞。当平均队列大小超过当前阈值时，网关会以一定的概率丢弃或标记每个到达的数据包。文献[10]提出了一种基于虚拟队列的标记方案，用于互联网路由器中的主动队列管理。该方案通过设置一个自适应虚拟队列，利用系统动力学的线性化模型，设计一个简单的规则来设计自适应虚拟队列算法的参数。文献[11]提出了一种 GEMINI 的设计与实现方案，该方案集成了 ECN 和延迟信号，用于跨直流拥塞控制。为了实现低延迟，GEMINI 使用延迟信号限制数据控制之间的延迟，并使用 ECN 防止数据控制内数据包丢失。然而，对于延迟敏感型应用程序，基于丢失信号的拥塞控制方法无法保证较低的传输时间。此外，数据包丢失并不总是由网络拥塞引起（例如随机丢包），这可能会误导拥塞控制决策。

基于延迟信号的拥塞控制方法是通过检测由网络引发的传输延迟进行调整。例如，文献[12]提出了一种高速 TCP 协议，是对 TCP 拥塞控制机制的一种修改，适用于具有大拥塞窗口的 TCP 连接。TCP Vegas 比大多数 TCP 变体提供了显著的性能提升，特别是在利用易出错链路的网络上使用时。然而，在研究延迟对 TCP 的影响时人们发现，当网络往返时间超过 50ms 时，TCP Vegas 的性能会显著下降。文献[8]对 TCP Vegas 存在的这种性能下降行为进行了研究与分析，提出了三种针对发送方进行修改的改进方案（称之为 TCP New Vegas）。与基于丢失信号的拥塞控制方法相比，基于延迟信号的拥塞控制方法在高动态场景（如无线网络）中更为适用，因为其不受随机丢包的影响。然而，精确测量传输延迟仍是一个主要挑战，例如在 Linux 协议栈中轻微的处理偏差就有可能导致测量误差，从而影响基于延迟的发送速率的精确控制。

混合拥塞控制方法是一种将基于丢失信号的拥塞控制方法和基于延迟信号的拥塞控制方法的技术优势结合起来的方法。例如，文献[13]提出了一种 TCP 方案—TCP Hybla，是对 TCP 标准版本 Tahoe、Reno、NewReno 进行修改，以消除对往返时间(Round trip time, RTT)的性能依赖。针对现有的 TCP 方案在大延迟的高速网络应用中存在严重的 RTT 不公平问题，文献[14]提出了一种旨在支持 TCP 友好性和带宽可扩展性的同时能减轻 RTT 不公平的拥塞控制方案。该方案使用了两种窗口大小控制策略（称之为加性增加策略和二分查找增加策略），大增量的加性增加策略确保了平方 RTT 的不公平性以及良好的可扩展性，而二分查找增加策略则在较小的拥塞窗口下支持 TCP 友好性。应该指出的是，混合拥塞控制方法依然难以通过与数据包丢失和传输延迟相关的隐式信号精确识别网络状态。

基于主动估算的 BBR 算法^[2]是基于带宽和 RTT 的主动估算来优化传输速率。BBR 通过持续探测网络瓶颈带宽和最小 RTT,跳脱了传统丢包驱动的限制性,在高带宽低延迟的网络中表现出色。但是, BBR 算法在多流竞争环境中容易导致资源分配不均,部分流量可能长期处于饥饿状态。此外,在动态变化剧烈的网络条件下, BBR 算法对网络状态的估算可能失准,从而导致性能波动。

随着网络环境的复杂性和多样性不断增加,传统的拥塞控制方法难以在多变的网络条件下保持稳定性能。这主要是因为传统方法通常依赖对网络条件的理想化假设,然而这些假设在实际网络中并不总是成立。即使在同一网络环境中,流量模式的动态变化也可能显著影响算法的稳定性和资源利用效率。在真实网络流量下,传统方法往往会出现性能下降和资源利用不足的问题^[15]。此外,现代网络架构通常包含有线网络和无线网络的组合,这进一步增加了数据包丢失检测的难度^[16]。因此,针对复杂网络环境,迫切需要一种能够自适应调整传输策略的智能拥塞控制方法。

3 基于强化学习的拥塞控制研究进展

为了解决上述问题,人们提出基于学习的拥塞控制算法。与传统方法不同,基于学习的拥塞控制算法通过训练模型使算法能够依据实时网络状态进行决策,而不是依赖预设规则,从而在动态和复杂的网络场景下表现出更强的适应能力。在拥塞控制任务中,基于学习的方法能够设定清晰且直接的优化目标,减少传统方法中因网络环境不确定性带来的局限。现有机器学习方案主要采用监督学习、无监督学习和强化学习三类方法。其中,监督学习和无监督学习已被广泛应用于网络状态估计,如拥塞信号和队列长度^[17]。然而,由于这些方法通常在离线环境下训练,它们无法在实际无线网络中准确区分随机丢包和拥塞丢包,从而限制了其实际应用效果^[18]。此外,大多数基于学习的拥塞控制算法主要通过调整拥塞窗口(CWND)来控制发送速率,而不是直接调整发送速率。在高速网络中,这可能导致突发性流量问题,即当多个 ACK 同时到达时,拥塞窗口可能会急剧增加,导致数据流瞬间过载,影响网络性能^[19]。

监督学习和无监督学习通常关注于即时状态的估计,缺乏对长期效果的考虑,而强化学习则更注重策略的长期回报^[20]。DRL 在应对真实网络环境的动态性和复杂状态空间时具有明显优势^[25],DRL 还能有效地从已有经典拥塞控制策略中学习知识^[31],以获得更优的拥塞控制策略,提高其对复杂网络环境的适应能力,并且所具备的在线学习能力对拥塞控制具有重要意义^[27]。目前,基于 DRL 的拥塞控制方法,根据其是否结合了传统的拥塞控制算法,可以分为混合方法

和全新设计方法(Clean-slate approaches)。下面分别介绍两类方法的研究进展情况。

3.1 混合方法

混合方法结合了传统的拥塞控制算法(如 TCP Cubic 或 BBR)或者传统的优化方法(如梯度优化或凸优化),这种结合可能仅限于训练阶段,也可能同时应用于训练和实际运行阶段。例如,文献^[4]提出的 Eagle 就是一个仅用于训练阶段的典型案例,它利用现有算法中的专家知识,并使用深度强化学习来训练一个广义模型。它结合应用的专家拥塞控制算法 BBR 来帮助训练 DRL 代理中的长短期记忆(LSTM)神经网络,希望能做出与专家一样好甚至更好的决策。

文献^[21]提出的 Orca 和文献^[22]提出的 Spine 是则同时应用于训练和实际运行阶段。这类方法通过训练 DRL 代理来优化传统拥塞控制算法(即专家)的决策过程,并采用双重控制流程:在高频事件(如基于 ACK 触发的拥塞窗口调整)上,由传统 CC 精细控制,确保实时性和稳定性;在低频决策(如周期性策略更新)上,由 DRL 代理执行,以优化长期传输策略,使系统更具适应性。

Orca 的设计初衷是整合传统 CC 算法与 DRL 技术,以同时发挥两者的优势。传统 CC 算法在已知网络环境下具有稳定性和低计算成本,但难以适应复杂多变的网络条件,而纯 DRL 方案尽管具备较强的自适应能力,但容易受限于训练数据的泛化性,且计算开销较高。为此,Orca 采用了双层控制架构,即细粒度控制由传统 CC 负责,确保对 ACK 响应的稳定性,而 DRL 代理执行粗粒度控制,以优化长期决策。其中,经典 TCP 算法用于实时调整拥塞窗口,而 DRL 代理则基于环境统计数据定期计算和执行新的拥塞窗口,使系统能够在动态网络环境下实现更优的传输策略,同时避免纯学习型方法的局限。

Spine 在 Orca 的双重控制架构基础上进一步优化,重点针对减少 DRL 计算开销进行了改进。在高吞吐量场景下,DRL 代理的高频决策可能占用大量 CPU 资源,影响数据路径性能。为了解决这一问题,Spine 采用了分层拥塞控制架构,将拥塞控制任务拆分为不同时间尺度上的两个子任务,并用不同组件分别处理。细粒度控制由轻量级 CC 执行器负责,以快速响应动态带宽变化,确保低延迟和高实时性,而粗粒度控制则由 DRL 代理负责,在较长时间尺度上生成控制子策略,交由 CC 执行器执行,以减少推理频率,从而降低计算成本。此外,Spine 通过监控模块动态调整 DRL 代理的决策频率,避免了固定时间间隔推理计算带来的额外开销,使系统能更有效地利用计算资源。

此外,文献[23]通过对现有的基于 DRL 的拥塞控制算法进行系统评估,总结了公平性、效率和稳定性等关键指标的重要性,并指出多目标优化框架在提升公平性和整体网络性能中的应用潜力。文献[24]进一步通过在线强化学习与帕累托优化的结合,提出了一种公平性优化算法,在多流竞争环境中显著改善了资源分配的公平性,而这一成果也被认为是多目标优化的典型案例。

3.2 全新设计方法

全新设计方法从训练一个不依赖专家的新策略开始。这种策略的改进完全依赖于代理积累的经验,由其自身的决策过程驱动。全新设计的拥塞控制算法的典型例子包括采用 Q 学习的算法^[25]、采用 Actor-Critic 方法的算法^[26],以及使用近端策略优化 (PPO) 的算法 (如 Aurora^[3])。

文献[27]开发了一种基于深度强化学习的无模型智能拥塞控制框架 (Drinc),在处理复杂多变的网络环境方面展现了巨大的潜力。Drinc 框架能够以一组测量特征的形式学习过去的经验,从而决定如何调整拥塞窗口大小。相应提出的拥塞控制方案无需依赖网络、调度或网络流量的精确模型,也不需要预先训练数据,在不同的网络条件下具有良好的鲁棒性。

为了能够根据特定的网络架构或应用程序修改核心拥塞控制方法来优化 TCP 性能,文献[28]将基于强化的 Q 学习框架与 TCP 设计结合起来,提出了一种称为 QTCP 的方法。该方法使发送方能够以在线方式逐渐学习最佳拥塞控制策略,不需要硬编码规则,因此可以推广到各种不同的网络场景。

文献[25]提出了一种基于学习的多路径拥塞控制方法 (称为 SmartCC),用于处理异构网络中多条通信路径的多样性。SmartCC 采用异步强化学习框架来学习一组拥塞规则,使发送方能够观察环境并采取行动自适应地调整子流的拥塞窗口,以适应不同的网络情况。为了解决高维空间中的无限状态问题,该文献作者还设计了一种用于状态聚合的分层瓦片编码算法和一种用于 Q 学习的函数估计方法,可以有效地推导出最优策略。仿真实验结果表明 SmartCC 能显著提高总吞吐量,并在各种性能指标上优于最先进的机制。

文献[26]提出了一种基于 DRL 的控制框架 (DRL-CC),利用单个 (而不是多个独立的) 代理对终端主机上的所有活动 MPTCP 流动态地联合执行拥塞控制,以最大化整体效用。该框架的新颖之处在于在 DRL 框架下利用灵活的递归神经网络来学习所有活动流的表示并动态处理它们,并将基于递归神经网络的表示网络集成到用于连续 (拥塞) 控制的演员-评论家 (actor-critic) 框架中,该框架利用新的确定性策略梯度以端

到端的方式训练评论家 (critic)、演员 (actor) 和递归神经网络。

文献[3]提出的 Aurora 算法首次强调了决策时间间隔在强化学习拥塞控制中的关键作用,通过动态调整决策间隔来有效平衡算法的实时性和计算开销。这一创新思路为后续研究提供了重要的参考,尤其是在动态网络条件下,通过适应性调整决策频率,显著提升了系统的性能稳定性与资源利用效率。

4 面临的挑战与研究思路

4.1 面临的挑战

随着云应用等对更高吞吐量和更低延迟的强烈需求,人们对拥塞控制的研究兴趣与日俱增。当多个客户端以不协调和分散的方式共享网络时,拥塞控制算法的设计尤其具有挑战性。网络中的每个客户端为了解网络状态所能进行的观测都是有限的。在大多数情况下,客户端无法获得共享网络的其他客户端的任何信息,例如它们的需求、交付率、往返时间 (RTT) 或丢失率。尽管可视性有限,但如何实现公平和最佳的资源共享、减少学习决策中的高计算成本是我们面临的基本挑战。

经过研究分析,我们认为目前流行的基于强化学习的拥塞控制算法仍然存在一些问题。首先,这些算法往往忽略了明确考虑公平性的必要性,在与其他流量共享同一瓶颈网络带宽时,Aurora 和 Eagle 等算法并不鼓励公平行为^[24]。此外,Orca 和 TCP-Drinc 等众多算法在其环境状态或奖励函数均未显式包含任何与公平性相关的信号。对 Orca 而言,使用对公平性相对友好的 AIMD 作为细颗粒度的控制方法,仍对公平性没有多大帮助。我们认为,在设计新的拥塞控制算法时,公平性应该是一个重要的考虑目标,也是现有基于 RL 的 CC 方法的主要局限。

其次,当前众多的 TCP 设计需求以及大量现有的 CC 算法表明,现有的 CC 算法要么是“样样通却不精”

(即泛化能力有限),要么是“聚焦型解决方案”,仅在某些特定网络特性下表现良好,但在其他网络环境中会失效。因为传统的 CC 设计策略通常依赖于对网络的某些假设,当这些假设不成立时,性能会显著下降。也就是说,我们仍然缺乏一种能够在不同环境中保持一致高性能且无需针对每个环境进行手动调优的自适应 TCP 算法。尽管基于 DRL 的 CC 方案前景广阔,但其发送速率调整所需的复杂模型推理却带来了高计算开销。这种高推理开销会消耗不可忽略的 CPU 资源。现有的解决方案通过降低推理频率来延长响应间隔,以此来解决开销问题。在间隔期间,CC 不受 DRL 控制,要么不进行速率调整^{[3][29]},要么依靠经典方案 (如 Orca 中的 Cubic 方案) 作为补救。因此,

它们无法充分利用 DRL 模型提供的性能优势,而且由于粗粒度控制,很容易受到网络拥塞的影响^[30]。在复杂的网络情况下,固定时间间隔的采样并不能很好捕捉到网络情况的变化。因此,我们思考一个问题:基于 DRL 的 CC 能否为每个 ACK 找到合适的时间控制间隔,以提供较为精细的颗粒度,在保证性能的同时下保持较低的计算开销。

此外,目前以 Aurora 为代表的基于强化学习的拥塞控制算法,虽然能够实时自适应地调整发送速率,但其决策过程依赖复杂的模型推理,导致计算开销较高,给实际部署带来了困难。同时,这些方法通常缺乏对公平性的明确优化目标,例如 Eagle 等算法并未在训练时显式考虑多流场景下的公平性问题,可能导致在实际运行中各个流量之间资源分配不均衡,进而降低整体网络性能。复杂的模型结构和繁重的训练过程进一步提高了实际部署的门槛。尤其是在高动态的网络环境下,现有方法的收敛速度偏慢,参数调优困难,显著影响了在真实网络中高效、稳定的应用。

4.2 我们的研究思路

基于2.1节的分析,为了解决多链路情况下共享同一网络链路时会造成资源分配不均,进而影响整体网络性能的问题,我们在尽可能保证链路效率的情况下,构建一种多目标优化的拥塞控制策略。在该策略中,将公平性显式的纳入奖励函数中,提升在复杂多链路竞争环境下的整体性能,通过动态的采样间隔设计以及采用双精度的控制方法,减少系统在运行过程中因学习和监测网络状况带来的消耗,最后对公平性和采样间隔以及网络性能建模为一个多目标优化问题,通过优化问题的求解,提升网络整体的性能效率。通过研究,将进一步推动基于强化学习的拥塞控制方法在实际网络中的应用,为复杂网络环境下的公平、高效数据传输提供新的解决方案。

为此,拟将开展以下方面的研究:

(1) 进一步扩展双重控制机制的应用,构建一个拥塞控制模型。该模型以强化学习和传统的混合拥塞控制算法为基础,建立一个双精度控制的体系,由传统拥塞控制算法执行细颗粒度的控制,由强化学习的拥塞控制算法执行粗颗粒度的控制,并定期由强化学习的拥塞控制去生成策略更新传统拥塞控制算法中的参数。通过引入强化学习模型对时间间隔进行自适应优化,使得时间间隔的调整能够有效平衡资源利用与性能优化之间的关系。低频决策能够根据网络状态的变化实时调整更新频率,从而在减少资源开销的同时,进一步提升系统性能的稳定性和适应性。

(2) 结合可变采样间隔双精度的DRL CC算法,将公平性作为指标显式地加入强化学习的奖励函数中,

并且以多目标优化的思想,将采样时间间隔也加入优化目标,以降低采样带来的资源消耗,从而研究一种多目标优化的拥塞控制优化策略。该策略结合网络性能、网络公平性、系统资源消耗等性能因素,建立目标函数,通过多目标优化,确保在动态环境下实现多性能目标(如公平性与效率)的协调优化,提升多链路环境下的整体性能,最大权衡系统中的链路性能,实现整体系统的性能最大化。

参考文献

- [1] Ha S, Rhee I, Xu L. CUBIC: A New TCP-Friendly High-Speed TCP Variant[J]. *Operating systems review*, 2008, 42(5): 64-74.
- [2] Cardwell N, Cheng Y, Gunn C S, et al. BBR: Congestion-based congestion control[J]. *Communications of the ACM*, 2017, 60(2): 58-66.
- [3] Jay N, Rotman N, Godfrey B, et al. A deep reinforcement learning perspective on internet congestion control[C]// *International Conference on Machine Learning*. PMLR, 2019: 3050-3059.
- [4] Emara S, Li B, Chen Y. Eagle: Refining congestion control by learning from the experts[C]// *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020: 676-685.
- [5] Nie X, Zhao Y, Li Z, et al. Dynamic TCP initial windows and congestion control schemes through reinforcement learning[J]. *IEEE Journal on Selected Areas in Communications*, 2019, 37(6): 1231-1247..
- [6] S Abbasloo, C Y Yen, H J Chao. Wanna make your TCP scheme great for cellular networks? let machines do it for you![J]. *IEEE Journal on Selected Areas in Communications*, 2021, 39(1): 265-279.
- [7] Floyd S, Henderson T, Gurtov A, et al. The newReno modification to TCP's fast recovery algorithm[EB/OL]. RFC 3782, 2004:1-18, <https://www.rfc-editor.org/rfc/rfc3782>
- [8] J Sing, B Soh. TCP New vegas: improving the performance of TCP vegas over high latency links[C]// *Fourth IEEE International Symposium on Network Computing and Applications*, IEEE Press, 2005:73-82.
- [9] S Floyd, V Jacobson. Random early detection gateways for congestion avoidance[J]. *IEEE/ACM Transactions on Networks*, 1993, 1 (4):397-413
- [10] S S Kunniyur, R Srikant. An adaptive virtual queue (AVQ) algorithm for active queue management[J]. *IEEE/ACM Transactions on Networking*, 2004, 12(2):123-134
- [11] G Zeng, W Bai, G Chen, et al. Congestion control for cross-datacenter networks[C]// *27th IEEE International Conference on Network Protocols*, Chicago, IL, USA, IEEE Press, 2019: 1-12
- [12] S. Floyd. Highspeed TCP for large congestion windows[R]. RFC 3649, 2003:1-34, <https://dl.acm.org/doi/pdf/10.17487/RFC3649>
- [13] C Caini, R Firrincieli. TCP Hybla: a TCP enhancement for heterogeneous networks[J]. *International Journal of Satellite Communications and Networking*. 2004, 22 (5): 547-566
- [14] L Xu, K Harfoush, I Rhee. Binary increase congestion control (BIC) for fast long-distance networks[C]// *The 23rd Annual Joint Conference of the IEEE Computer and Communications Societies*, Hong Kong, China, IEEE Press, 2004: 2514-2524

- [15] Dong M, Li Q, Zarchy D, et al. {PCC}: Re-architecting congestion control for consistent high performance[C]//12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15). 2015: 395-408.
- [16] R Mittal, N Dukkupati, E Blem, et al. TIMELY: RTT-based congestion control for the datacenter[J]. ACM SIGCOMM Computer Communication Review, 2015, 45(4):537-550
- [17] C P Fu, S C Liew. TCP Veno: TCP enhancement for transmission over wireless access networks[J]. IEEE Journal on Selected Areas in Communications, 2003, 21 (2): 216-228
- [18] R King, R G Baraniuk, R H Riedi. TCP-africa: an adaptive and fair rapid increase rule for scalable TCP[C]// The 24th Annual Joint Conference of the IEEE Computer and Communications Societies, Miami, FL, USA, IEEE Press, 2005: 1838-1848
- [19] S Ryu, C Rump, C Qiao. Advances in internet congestion control[J]. IEEE Communications Surveys & Tutorials. 2003, 5(1): 28-39
- [20] Li Y. Deep reinforcement learning: An overview[J]. arXiv preprint arXiv:1701.07274, 2017.
- [21] S Abbasloo, C Y Yen, H J Chao. Classic meets modern: A pragmatic learning-based congestion control for the internet[C]// 2020 Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication, ACM Press, 2020: 632-647.
- [22] H Tian X Liao, C Zeng, et al. Efficient DRL-based congestion control with ultra-low overhead[J]. IEEE/ACM Transactions on Networking, 2024, 32(3):1888 - 1903
- [23] L Giacomoni, G Parisi. Reinforcement learning-based congestion control: a systematic evaluation of fairness, efficiency and responsiveness[C]// 2024 IEEE Conference on Computer Communications, Vancouver, BC, Canada, IEEE Press, 2024: 1451-1460
- [24] S Emara, F Wang, B Li, et al. Pareto: Fair Congestion Control With Online Reinforcement Learning[J]. IEEE Transactions on Network Science and Engineering, 2022, 9(5): 3731-3748
- [25] W Li, H Zhang, S Gao, et al. SmartCC: A reinforcement learning approach for multipath TCP congestion control in heterogeneous networks[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(11):2621 - 2633
- [26] Z Xu, J Tang, C Yin, et al. Experience-driven congestion control: When multi-path TCP meets deep reinforcement learning[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(6): 1325 - 1336
- [27] K Xiao, S Mao, J K Tugnait. TCP-drinc: Smart congestion control based on deep reinforcement learning[J]. IEEE Access , 2019, 7: 11892-11904,
- [28] W Li, F Zhou, K R Chowdhury, et al. QTCP: Adaptive congestion control with reinforcement learning[J]. IEEE Transactions on Network Science and Engineering, 2019, 6(3): 445 - 458
- [29] Y Ma, H Tian, X Liao, et al. 2022. Multi-objective congestion control[C]// Proceedings of the 17th European Conference on Computer Systems. Rennes France , ACM Press, 2022:218-235.
- [30] Zhang J, Zeng C, Zhang H, et al. Liteflow: towards high-performance adaptive neural networks for kernel datapath[C]//Proceedings of the ACM SIGCOMM 2022 Conference. 2022: 414-427.
- [31] Yen C Y, Abbasloo S, Chao H J. Computers can learn from the heuristic designs and master internet congestion control[C]//Proceedings of the ACM SIGCOMM 2023 Conference. 2023: 255-274.