

用户密集环境下基于边缘智能的直播视频传输优化机制

顾晓丹, 吴文甲, 凌振

(东南大学计算机科学与工程学院, 江苏 南京 211189)

摘要: 针对传统直播视频传输优化机制部署在服务器侧, 无法快速响应用户终端所处的无线网络环境动态变化的问题, 提出基于边缘智能的直播视频传输优化机制 S-Edge。该机制部署在基于 OpenWrt 的无线接入点上, 综合利用占空比、信噪比等无线信道状态信息, 基于模糊逻辑理论对终端优先级及传输速率进行智能决策, 并通过分层令牌桶的主动队列管理和业务需求驱动的无线传输速率自适应控制技术, 实现直播视频终端业务的实时调度。为了验证所提机制 S-Edge 的有效性和性能, 在真实场景下搭建基于多射频接口的硬件实验平台并开展用户密集无线网络环境下的实验。实验结果表明, S-Edge 可以显著降低平均时延和丢包率, 有效提升用户密集环境下直播视频传输业务的服务质量指标。

关键词: 无线局域网; 直播视频传输; 用户密集; 边缘智能

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023232

Live video transmission optimization mechanism based on edge intelligence in high client-density environment

GU Xiaodan, WU Wenjia, LING Zhen

School of Computer Science and Engineering, Southeast University, Nanjing 211189, China

Abstract: The traditional live video transmission optimization mechanism is deployed on the server side, which cannot quickly respond to the dynamic changes of the user's wireless network environment. To address this problem, a live video transmission optimization mechanism based on edge intelligence called S-Edge was proposed. It was deployed on the OpenWrt-based wireless access point, and comprehensively utilized the wireless channel state information such as airtime utilization and signal-to-noise ratio to make intelligent decisions on terminal priority and transmission rate based on fuzzy logic theory. Furthermore, the active queue management with hierarchical token bucket and service demand-driven wireless transmission rate adaptive control technologies were introduced to realize the real-time scheduling of live video data. In order to verify the effectiveness and performance of the proposed mechanism, a high client-density environment was built through user clusters based on multi-radio interfaces in the real-world scenario. Experimental results show that S-Edge can significantly reduce the average delay and packet loss rate, which meets QoS requirements of live video transmission services in the high client-density environment.

Keywords: WLAN, live video transmission, high client-density, edge intelligence

0 引言

近年来, IEEE 802.11 无线局域网 (WLAN, wireless local area network) 技术快速发展, 传输速率越来越高, 并广泛应用于家庭影音、企业会议、

大型活动等诸多场景, 用以提供便捷、高速的无线网络接入方式。Cisco 白皮书指出, 2023 年 Wi-Fi 平均传输速率将增长到 110.4 Mbit/s, 为 2018 年的 2.2 倍。但随着终端设备数量的爆炸式增长, WLAN 大多处于用户密集的状态。再加上实时会议、在线

收稿日期: 2023-08-02; 修回日期: 2023-11-01

基金项目: 国家自然科学基金资助项目 (No.62072102, No.62132009, No.62102084)

Foundation Item: The National Natural Science Foundation of China (No.62072102, No.62132009, No.62102084)

教育、电商直播等应用的快速普及,使用户密集型无线网络环境下的直播视频业务越来越普遍。考虑到用户密集型环境下,业务种类繁多,非实时业务类产生的流量对直播视频流量传输造成严重干扰,同时在多终端争夺共享带宽的瓶颈网络环境中,终端吞吐量将会出现较大波动,进而造成用户体验(QoE, quality of experience)下降。因此,尽管无线网络的传输带宽越来越大,但直播视频传输业务的QoE依然较差,亟待优化。

实时的视频传输主要可以分为点播和直播两类业务,其中,点播业务倾向于充分利用网络资源传输高码率视频提升QoE;直播业务更注重流畅度,码率决策倾向于低码率视频,用以缓解网络波动带来的服务质量(QoS, quality of service)指标波动。因此对于视频直播而言,播放时延和视频中断次数对QoE影响更为严重,再加上用户密集环境下高负载接入点(AP, access point)上的业务通常会经历高时延和高丢包率^[1],无法满足低时延需求,导致QoE急剧下降。对此,研究人员提出对终端和服务器之间的网络状态进行细粒度测量从而实现视频编码的动态调整^[2-4]。边缘智能的迅猛发展也为直播视频优化提供了一种新的思路^[5-7],即借助边缘计算节点优先响应处理视频任务或借助深度学习技术增强视频质量。然而,上述大部分工作均未考虑最后一跳即边缘AP的影响,导致在用户密集无线网络环境下直播视频用户体验不佳;且忽视了低优先级业务诸如浏览网页、下载文件等带来的影响。此外,现有研究工作^[5-6,8]往往都通过仿真实验对提出的算法或机制加以验证,没有在真实环境中进行实验,其结果可能与实际性能存在较大差距。

针对上述问题,本文提出了一种基于边缘智能的直播视频传输优化机制,主要研究工作如下。

1) 针对用户密集无线网络环境下终端间竞争带宽造成直播业务QoS波动明显的问题,提出一种基于AP边缘智能的直播视频传输优化机制S-Edge,综合考虑终端信道状态,利用模糊逻辑理论对终端优先级及传输速率进行决策,实现对直播视频流量的实时优化调度,确保其在高负载AP上的QoS指标平稳,从而提高直播业务的用户体验。

2) 在真实场景下,搭建基于多射频接口的硬件实验平台来构造用户密集无线网络环境,在降低部署成本的前提下实现对相关无线信道状态的获取;

设计两套实验方案对所提优化机制的有效性和性能进行了验证,并与现有直播视频优化机制LoL+^[4]进行对比,实验结果表明,S-Edge可以显著降低时延和丢包率,能够满足用户密集环境下直播视频传输业务的QoS需求。

1 相关工作

直播视频业务场景下传输优化的主要目的是维持客户端与服务器之间的网络连接稳定并控制时延在一个可接受的范围内。为了维持连接稳定,基本解决思路是对终端和服务器之间的网络状态进行细粒度测量,从而动态调整编码器产生的视频文件大小,进而适配网络环境,避免网络波动造成的直播视频卡顿。Luo等^[7]指出当前带宽受限的设备无法支持高清直播视频流,因此基于神经网络视频增强技术,提出高清直播流框架LiveSR,该框架可以使用较低的带宽支撑高质量的直播视频流,但需要定制编码器。Hsieh等^[1]研究直播视频中的中断和播放时延两者之间的关系,指出在高负载无线网络环境下视频中断和播放时延成反比,因此可以通过适当提升用户的播放时延以降低视频中断的发生。但该文并未考虑无线网络上其他业务的存在对直播视频传输的影响,且该调度策略仅在仿真环境下进行测试。Zhang等^[2]为了解决模拟环境与实际环境之间差异较大的问题,提出了一种基于在线强化学习的解决方案,通过在每个终端上部署强化学习模型,而后将它们聚合产生一个新的强化学习模型,再对每个终端进行更新,可以使终端快速适应其未经历过的网络状态。在此基础上,该文还将在在线强化学习与已有的决策算法相结合,避免了强化学习在探索阶段造成的QoE下降问题。类似地,胡雨涵^[9]在联邦学习模型聚合算法和模型自适应算法的基础上,提出了Chamelone算法,解决了在线学习需要大量服务器资源的缺陷。但实际环境中网络波动往往发生在最后一跳即AP上,此类方案均需要对客户端修改。Zhang等^[3]分析了直播内容与观看用户数量之间的关系,利用长短期记忆(LSTM, long short-term memory)网络对电子竞技中的游戏流程进行预测,结合游戏流程中不同状态与用户观看人数的关系以及不同规格云计算服务器编码视频的速度,提出了在服务端根据直播内容动态调整编码能力以降低编码时延进而优化直播时延的方案。但该方案缺乏泛化性,对电子竞技这类流程明确的直播内容有一定

的效果。Bentaleb 等^[4]提出基于 HTTP 的动态自适应流 (DASH, dynamic adaptive streaming over HTTP) 的直播方案尽管提供了多种码率供选择,但自适应决策算法缺乏相应的学习性,故基于自组织映射 (SOM, self organizing map) 方法提出一种 LoL+增强机制,利用启发式的码率决策算法,通过综合考虑目标时延与当前时延的差距,确定播放速率与码率,用以降低端到端的时延。Gutterman 等^[10]基于滑动窗口对带宽和时延进行估计,在此基础上部署丢帧机制,即对无法满足时延要求的视频块进行丢弃重新缓冲;并部署播放速率控制机制,在不影响用户体验的情况下,若当前时延大于目标时延,将会调高播放速率。乔春雨^[11]针对真实场景下不同用户的差异,提出参数化的用户体验多样性模型,并设计相应的传输优化算法,以提高个性化的用户体验。

此外,研究人员还针对无线信道易受干扰导致视频业务 QoS 下降的问题,提出了一系列的解决方案。Ali 等^[8]指出随着无线密集型环境越来越普遍,需要一种普适的方法解决多用户间的竞争问题,因此基于深度强化学习 (DRL, deep reinforcement learning) 设计了一种 MAC 层的接入控制机制,通过获取当前终端的信道条件及其他状态信息,决策出恰当的退避窗口大小用以提高整体终端间吞吐量、接入时延及公平性。Bhattacharyya 等^[12]基于强化学习在 AP 上设计了一种调度系统 QFlow,该系统通过收集终端播放状态信息来决策视频数据的优先级,进而提升视频业务的用户体验。Babu 等^[13]指出在无线网络中传输视频受到无线介质特性、多径路由限制等因素影响,因此在保证 QoE 的基础上提出视频源的多径传输机制,利用自适应可扩展视频编码和中间路由节点的选择,实现高质量低时延的视频传输。

在实验验证方面,现有工作主要借助仿真软件和真实场景 2 种方式来构建用户密集环境。由于构建真实的用户密集场景成本较高、实施过程较为复杂,因此大部分的工作都采用仿真的方式。通常是作者自己构建仿真代码进行实现^[8,14],主要通过收集带宽信息,并利用 TC (traffic control) 构建模拟器^[15],编写特定的代码和应用实现终端的仿真,因此可以实现相同网络环境下的测试,但无法模拟实际环境中相邻节点的干扰和噪声。也有一些工作基于 NS-3^[16]或其他网络仿真工具进行设计。

总体来说,在直播视频传输业务场景中,大部

分工作未考虑最后一跳即 AP 的影响,导致在用户密集环境下直播视频用户体验较差;且忽视了低优先级业务诸如浏览网页、下载文件等带来的影响。此外,一些工作提出利用信道状态信息优化无线网络下的视频传输,但往往都使用仿真实验,缺乏真实场景下用户密集环境的实验验证。考虑到使用真实设备搭建用户密集环境成本耗费过高,本文选择搭建基于多射频接口的硬件实验平台^[17-19],在真实环境和仿真软件之间进行平衡,既可以获取相关无线信道状态,也可以降低部署成本。

2 基于边缘智能的直播视频传输优化机制

针对用户密集型无线网络场景下高负载 AP 上多终端争夺共享带宽导致直播业务用户体验较差的问题,本文提出了一种基于边缘智能的直播视频传输优化机制 S-Edge,通过在边缘 AP 感知终端无线信道状态,并利用模糊逻辑理论对终端优先级及传输速率进行决策,实现对直播视频流量的实时优化调度,从而提高直播业务的用户体验。

2.1 问题分析

随着网络业务的日益复杂以及边缘计算的飞速发展,研究人员引入了边缘智能的概念,将对直播视频的传输优化逐渐转移至边缘节点^[5-7],如在 WLAN 中的 AP 侧进行优化机制的部署。然而此类工作都忽视了 AP 可能带来的影响,尤其是在用户密集环境下,多终端间竞争信道使用时间,将会极大地增加报文等待时延,并伴随着过高的丢包率,导致直播视频用户的服务质量急剧下降。针对这一问题,本文考虑从以下 3 个方面进行优化。

1) 业务优先级

在用户密集场景中,终端承载了多种上层业务,不同的业务类型对传输的需求也不尽相同,既有对带宽需求极大但时延不敏感的文件下载类业务,也有可容忍一定丢包但时延、带宽敏感的实时类业务等。IEEE 802.11e 协议对不同业务的数据传输优先级进行定义,将数据流量被划分为以下 4 类。

① 语音服务 (Voice, AC_VO): 一般为 VoIP 流量类型,对时延最为敏感,优先级最高。

② 视频服务 (Video, AC_VI): 视频流量优先级低于语音服务。

③ 尽力传输 (Best-effort, AC_BE): 一般默认为尽力传输,例如网页访问的数据。

④ 背景流量 (Background, AC_BK): 对时延

要求最不敏感的流量，例如心跳通信包。

根据 IEEE 802.11e 标准的定义，直播视频传输这种类似于语音服务的业务应当处于高优先级。然而由于当前商业 AP 上大都不支持 IEEE 802.11e 标准，且用户需求的变化会改变 AP 中的流量分布，因此需要在 AP 上引入低开销的流量识别方法，从而为制定合理的调度策略提供决策依据。

2) 队列管理

现有的商用无线路由设备通常是集 AP 与路由器功能于一身，每个端口上存在一个或多个队列，用于处理数据包的转发及 QoS 保证功能。队列的重要作用在于吸收网络中突发的数据流量，较大的队列可以吸收更多的突发流量，提高吞吐量，但相应地会带来较高的排队时延。因此路由设备中会部署队列管理机制，通常通过丢包来管理队列长度，使其处于低时延高吞吐量的状态。其中，被动队列管理机制包括去尾、随机丢弃和从前丢弃 3 类，都是在队列满时触发，而主动队列管理机制则是在队列满前触发丢包，但依然需要处理排队时延和吞吐量之间的关系。除了队列管理，路由设备还需要设计合适的队列调度算法，结合相应优化目标设定优先级并进行调度，即决定下一帧发送哪个包。由于目前 IP 网络对 QoS 的考虑还未成熟，路由设备出于公平性原则，一般会为每条流设置相同的优先级。因此，对于直播视频流，在无线路由上现有架构缺乏相应的优化机制，且密集型无线网络环境下 AP 往往处于过载状态，需要对队列管理和调度算法进行改进，降低时延和丢包率，以满足直播视频的流畅度需求。

3) 无线传输速率调控

由于用户的移动性和无线信道的不稳定性，IEEE 802.11 协议组设计了用于控制当前发送速率的机制即传输速率调控机制，该机制控制调制与编码策略索引 (MCS index, modulation and coding scheme index) 用以确定无线帧的调制速率。通过该机制，可以有效地提高无线终端的吞吐量，降低重传率。以开源无线网卡驱动 ath10k 使用的 Minstrel 算法^[20]为例，发送方将发送过程分为采样部分和非采样部分。在采样部分，从预先设定的速率组中随机选择一个可用的速率，若可以获得更高的吞吐量，则使用该速率进行下一帧的传输，否则仍使用先前的传输速率；在进入非采样部分前，会统计出 3 个数据速率用以对应最佳吞吐量、次佳吞吐量和最大可能吞吐量。在非采样部分，将会一直

使用最佳吞吐量对应的速率进行传输，直到超过了最大重传次数后使用次佳吞吐量对应的传输速率，一定的概率使用最大可能吞吐量对应的传输速率。MCS index 控制机制是工作在无线 MAC 层的速率自适应方法，无论上层如何调控发送速度，理论上最大的发送速率的瓶颈由 MCS index 决定。由此可知，目前广泛部署的无线传输速率调控机制是上层业务无关的，可能会使一些时延不敏感的业务占据了大量的带宽，导致时延敏感的直播业务服务质量下降。因此可以考虑结合直播视频的传输需求，设计无线传输速率自适应控制方法，调控发送速率降低在无线传输层的丢包，并维持直播业务所需的带宽。

在上述分析的基础上，本文提出了一种基于边缘智能的直播视频传输优化机制 S-Edge，包括基于模糊逻辑的队列优先级调度方法、基于分层令牌桶 (HTB, hierarchical token bucket) 的主动队列管理方法以及业务需求驱动的无线传输速率自适应控制方法三部分，分别从队列优先级、队列管理和无线传输速率调控三方面进行改进优化，以降低丢包率，最大限度满足直播视频传输的流畅度需求。具体流程如图 1 所示，对于未知业务，首先进行业务判定，若为非直播业务，则为其构建低优先级队列；若为直播业务，则设定较高的处理能力，在此基础上，感知当前接入终端信道状态、AP 状态信息等，计算相应的 QoS 指标作为模糊逻辑系统的输入进行模糊逻辑决策，最后根据决策结果进行队列优先级和无线传输速率的调整。

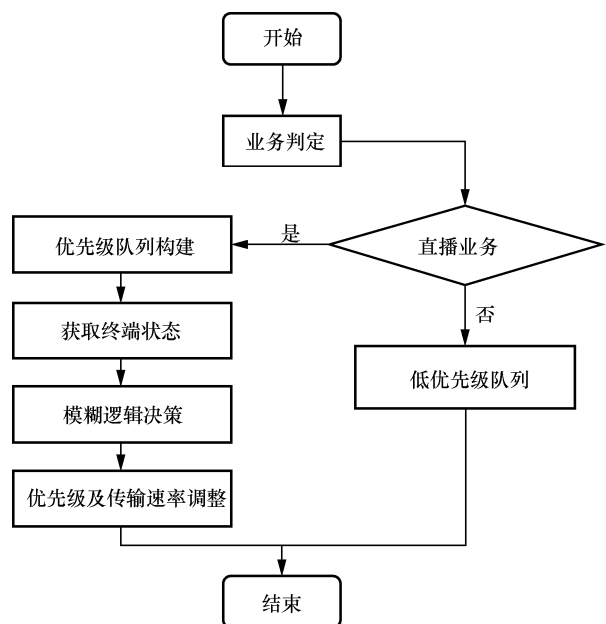


图 1 S-Edge 实现流程

2.2 基于模糊逻辑的队列优先级调度

在用户密集型无线网络环境中，多终端高度异构，所承载的业务类型繁多，且无线信道易受干扰，因此难以建立精确的数学模型来优化直播视频传输。针对该问题，本文引入模糊逻辑理论的思想，结合业务优先级和直播视频 QoS 指标构造模糊逻辑输入，依据设计规则构造模糊规则，然后进行模糊推理输出当前终端的优先级及对应的传输速率。

1) 应用流量分类

不同类型的业务对网络的 QoS 指标需求不同，相较于其他业务类型，直播视频传输更倾向于一个稳定的网络环境。目前，度量网络性能指标有带宽、时延、抖动和丢包率等，而主流的服务模型有尽力而为服务模型、集成服务模型和差分服务模型。其中差分服务模型可根据网络中流量的特征区分不同的类别并标记成不同的优先级。当网络发生拥塞时，对于不同类型的流会根据其所属优先级进行差分处理，同类的业务会被聚合在一起统一发送。因此，本文首先需要对上层的业务类型进行分类，然后利用差分服务模型进行优先级队列的调度。虽然在 TCP/IP 模型中各个层都可以做流量标记，但考虑到传输层、应用层识别规则较复杂，而无线 MAC 层设计规则识别的开销过大，因此本文采用了基于 IP 层报文头部规则匹配的分类方法，具体可通过采集大量的 IP 流量数据进行分析，导出流量匹配规则，然后利用 iptable 实现流量分类。对于非直播类型业务，构建低优先级队列；对于直播视频传输业务，则设定较高的处理能力。

2) 模糊逻辑系统

模糊逻辑控制^[21]理论是一种模仿人思维的模糊理论，其作为智能控制的一种，在实际应用中取

得了不错的效果，一般用于解决难以建立精准数学模型的问题。如图 2 所示，模糊逻辑系统包含模糊化、模糊推理机和解模糊 3 个部分。

①模糊化。模糊化的作用在于将输入精确量转换成模糊量，并映射至预先定义的模糊集合上。对于每个输入的语言变量，其取值作为一组语言名称，并将其所有取值构成语言名称集合，每个语言名称集合对应一个模糊集合。本文将输入语言变量定义为吞吐量 (Th)、丢包率 (Ls)、信噪比 (Sr)、空分流 (Ns) 和占空比 (Au) 5 种。其中，空分流指利用多个天线进行数据流传输的通道数，而占空比是指无线信道在给定周期内处于繁忙状态的时间百分比。本文将每种语言变量取值的模糊集合个数设为 3 个，即高、中和低，分别用 High (H)、Medium (M) 和 Low (L) 表示。在将输入值映射为模糊集合时需要确定隶属函数，普遍使用的隶属度函数包括三角形隶属度函数、梯形隶属度函数、钟形隶属度函数、Sigmoid 型隶属度函数和高斯型隶属度函数等。考虑部分输入语言变量例如吞吐量并无明确的区分度，故本文采用三角形隶属度函数，定义式如式(1)所示。

$$f(x,a,b,c) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ \frac{c-x}{c-b}, & b \leq x < c \\ 0, & x \geq c \end{cases} \quad (1)$$

其中，参数 a 和 c 确定三角形的“脚”， b 确定三角形的“峰”。本文结合专家经验和最小模糊度法分别确定了 5 种输入变量的隶属度函数，如图 3 所示。根据图 3 可以获得各个输入变量的模糊集合。例如图 3(a)为吞吐量 (Th) 对 H、M 和 L 这 3 个模糊标记的隶属度函数，其中 x 轴为 Th 的输入值，其输

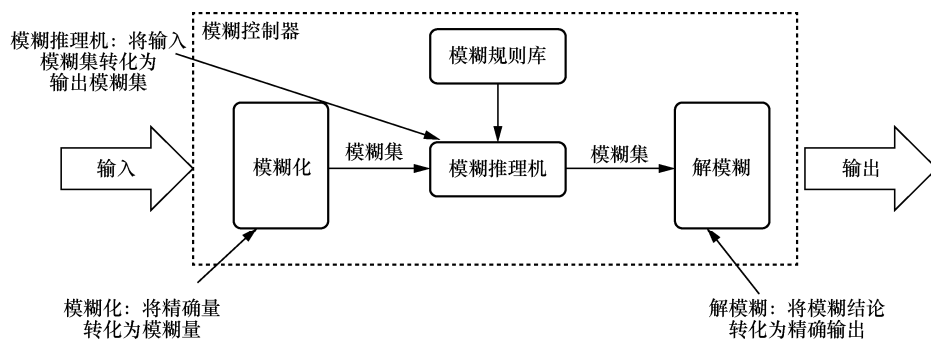


图 2 模糊逻辑系统结构

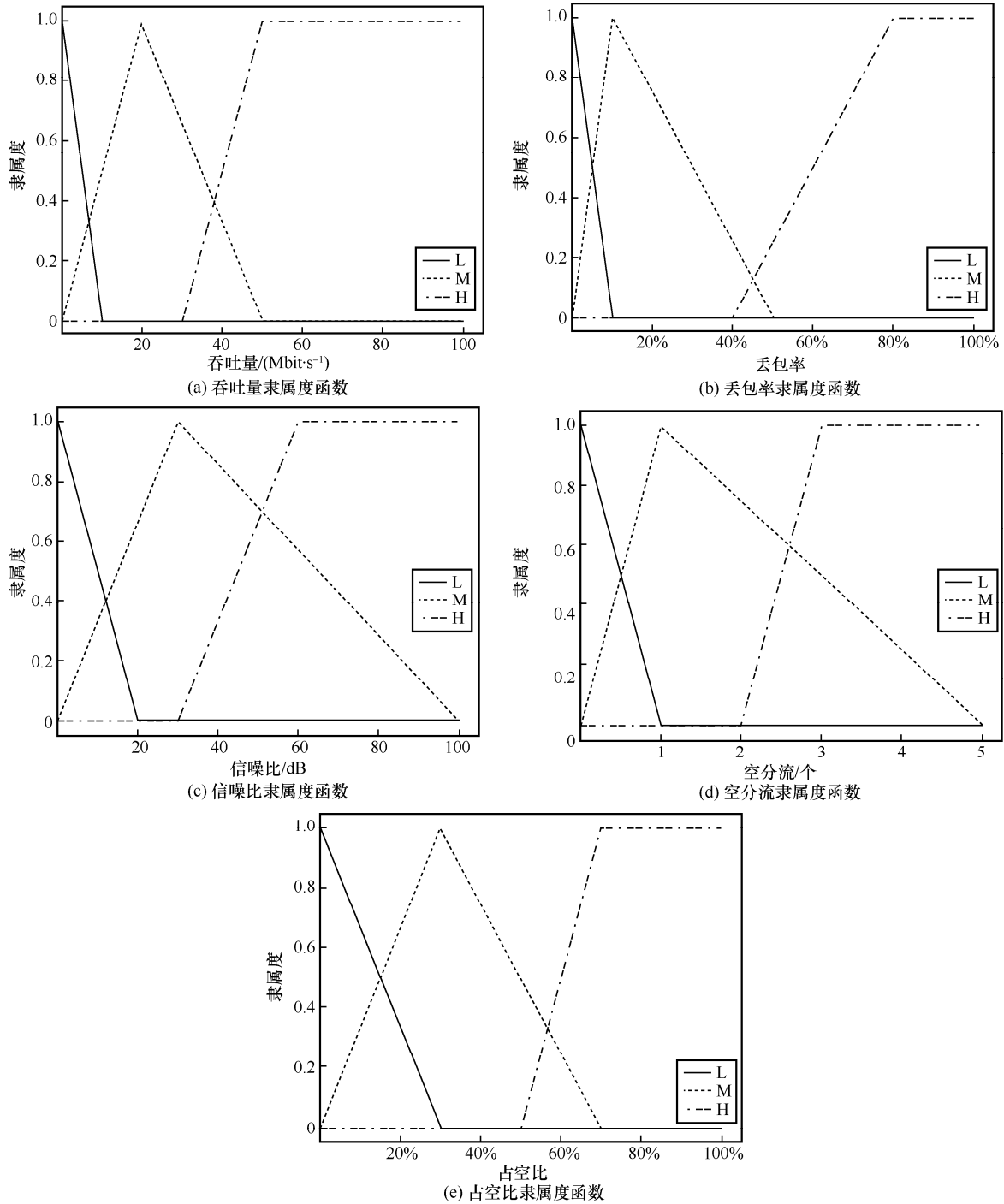


图 3 各输入变量的隶属度函数

出的模糊集合由 3 条折线 (H、M 和 L) 上对应点的 y 值构成。例如, 当 $Th=20$ Mbit/s 时, 其对于 H 的隶属度值为 0, 对于 M 的隶属度为 1, 对于 L 的隶属度值为 0。

② 模糊推理机。模糊规则库由一系列模糊规则组成, 本文采用 IF (满足一组条件) THEN (输出一组结果) 的形式, 在专家经验的指导下进行模

糊规则的制定。“IF-THEN”规则中的前提条件和结论均是模糊概念, 判断条件为各输入语言变量的模糊集合的组合, 输出为运行直播视频的各个终端的优先级 Pr 以及对应的传输速率 MCS index 的模糊集合。表 1 为制定的部分模糊规则示例, 其中第一条规则表示, 当吞吐量为低、丢包率为低、信噪比为中、空分流为中、占空比为高时, 则采取的策略为

低优先级、中传输速率。下面以 2 个输入一个输出的模糊规则为例进行说明。假设有 2 条模糊规则 R_1 和 R_2 ，具体定义如下。

R1: IF x is A_1 and y is B_1 THEN z is C_1

R2: IF x is A_2 and y is B_2 THEN z is C_2

表 1 模糊规则示例

规则	模糊规则条件 (IF)					模糊规则结果 (THEN)	
	Th	Ls	Sr	Ns	Au	Pr	MCS index
R_1	L	L	M	M	H	L	M
R_2	L	H	L	L	M	H	L
R_3	M	H	L	L	M	M	L
R_4	H	L	H	H	L	M	H
R_5	H	L	H	H	L	M	M

依据模糊逻辑运算规则，定义规则 R_1 和 R_2 的点火强度分别为 α_1 和 α_2 ，这里点火强度可用来评估该规则被激活的强度，计算方法如式(2)和式(3)所示。

$$\alpha_1 = \mu_{A_1}(x) \wedge \mu_{B_1}(y) \quad (2)$$

$$\alpha_2 = \mu_{A_2}(x) \wedge \mu_{B_2}(y) \quad (3)$$

其中， μ 是隶属度函数， \wedge 为模糊 AND 运算。模糊 AND 运算通常为取小运算或代数积，如式(4)所示，本文采用取最小运算。OR 运算采用取最大运算，NOT 运算取反，即用 1 减去。

$$\alpha_i = \mu_{A_i}(x) \wedge \mu_{B_i}(y) = \begin{cases} \min(\mu_{A_i}(x), \mu_{B_i}(y)) \\ \mu_{A_i}(x)\mu_{B_i}(y) \end{cases} \quad (4)$$

具体来说，对于表 1 中的第一条规则，首先根据隶属度函数获得 5 种输入变量分别对于 L、M、H 的隶属度值，即可知 Th 对于 L、Ls 对于 L、Sr 对于 M、Ns 对于 M 以及 Au 对于 H 的隶属度值；然后根据最小运算取 5 个值中的最小隶属度值作为该规则的点火强度。至此，对于每一条规则 R_i ，都可以计算其点火强度 α_i ，并将其作为该规则推理结果的模糊集合的隶属度值。例如，若根据最小运算获得表 1 中规则 R_1 的点火强度为 0.3，则意味着 Pr 对于 L 的隶属度值为 0.3，MCS 对于 M 的隶属度值为 0.3。

③ 解模糊。为了将模糊推理得到的模糊量转换为精确的数值，需要进行解模糊操作。本文采用最大隶属度方法进行解模糊，即将所有模糊规则推理结果的模糊集合中取隶属度最大的元素作为输出值，就可以有效获取优先级及传输速率的精确值。

2.3 基于 HTB 的主动队列管理

在用户密集型无线网络环境下，AP 处于高负载状态，随着用户的增加，直播视频传输业务的丢包率会逐渐加大。因此本文提出基于 HTB 的主动队列管理方法，实时调度高丢包率终端的优先级，以降低其被丢包的概率。

队列管理旨在维护路由器缓存队列长度，根据设定的策略应对拥塞发生后如何进行丢包。与被动队列管理相比，主动队列管理将会在缓冲队列未超前，通过当前的信息提前判断网络发生拥塞的概率，若判断会发生拥塞，则以一定概率丢弃数据包。因此在差分服务的基础上，考虑到网络的动态性及直播视频的流畅度需求，当存在非直播业务时，优先丢弃非直播业务数据包；对于直播业务，依据信道状态设定丢包权重，如对于低信道质量终端，给予低权重系数，由队列根据优先级自动丢包。本文采用了 HTB 算法构建树，利用分层控制思想对各节点的流量数据进行管理，并设定节点的优先级，且叶节点具备向中间节点借用带宽的功能，即根据模糊逻辑决策出的终端优先级将流量分配至不同的队列。

HTB 算法首先划分多个根节点、中间节点和叶节点，每个节点都作为一个类。根节点是一种广义上的分类，所有终端的数据包都会流入该节点。中间节点是划分叶节点的载体，可以包含多个叶节点管理规则，同时叶节点的带宽借用来自中间节点。叶节点是具体的执行规则，包含自己的数据包队列及带宽分配和优先级，如图 4 所示。其中每个类应用 HTB_CAN_SEND、HTB_MAY_BORROW 和 HTB_CANT_SEND 这 3 种模式，分别对应该类数据可以发送、借用父类带宽可以发送和该类数据不能发送。若同时有几个子类节点向父类节点借用带宽时，会优先满足优先级较高的节点，若几个节点优先级相同则均分剩余可用带宽。具体流程如图 5 所示，对于识别的直播业务流量，首先根据分层机制建立 HTB 树，树的每个节点作为一个类，每个叶节点上有一个子队列。对根节点、中间节点、叶节点赋予定义的可借用带宽因子、速率、最高速率，并在叶节点上定义优先级。AP 发出的数据包由各队列上的数据构成，其中高优先级队列优先发送。HTB 算法从树的底部开始往上找 CAN_SEND 状态的节点，如果找到某一层有 CAN_SEND 状态的节点则停止。

若该层是叶节点，则按照优先级排序，优先级高的先发送。若该层为中间节点，则顺着树往下找子孙叶节点；如果只有一个叶节点处于 MAY_BORROW 的状态，就将中间节点的可借用带宽因子值借给该叶节点，让其发出数据包；如果中间节点有多个子孙叶节点处于 MAY_BORROW 状态，则按优先级排序，优先级高的叶节点先发送。

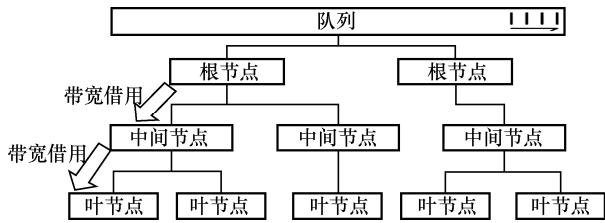


图 4 HTB 分层示意

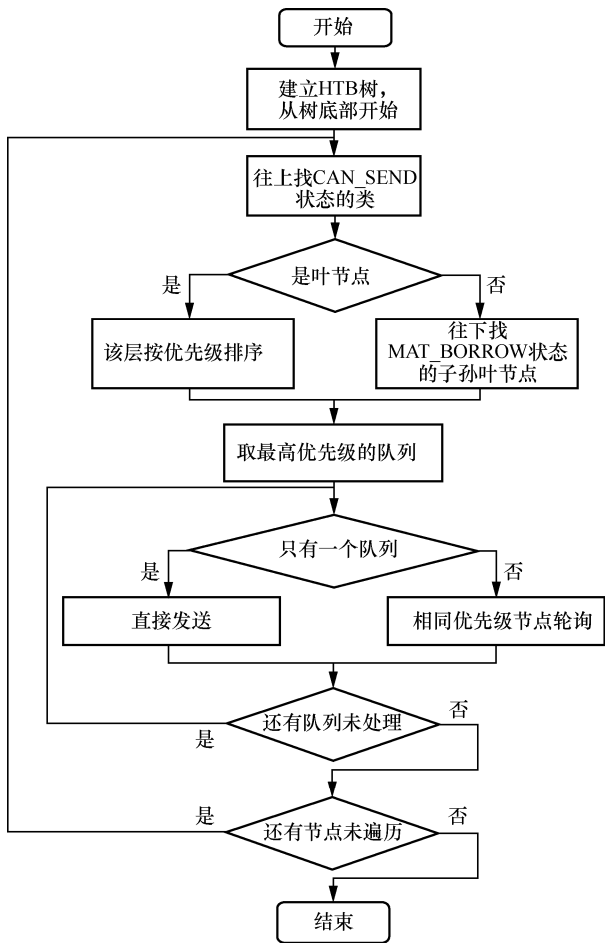


图 5 基于 HTB 的主动队列管理方法的具体流程

2.4 业务需求驱动的无线传输速率自适应控制

IEEE 802.11n 射频速率的配置通过 MCS index 实现。MCS 将所关注的影响通信速率的因素

作为列，将 MCS index 作为行，形成一张速率表。所以，每一个 MCS index 其实对应了一组参数下的物理传输速率。当前 Linux 无线子系统存在基于 Minstrel 算法的无线传输速率调整机制，主要控制数据帧的传输速率并最大化终端无线层面吞吐量。但在用户密集型环境中终端的可用无线信道由信道争用机制决定，其发送速率与上层业务需求存在不匹配问题。对于低时延低带宽需求的实时聊天信息，动态调整会适当降低其 MCS index 和解码错误概率，但 MCS index 的降低会带来传输时间的增加，将会加剧用户密集型环境下信道的争用，导致时延增加。因此，本文对现有的无线传输速率调整机制进行改进，面向直播业务需求，设计一种自适应的 MCS index 主动控制方法，并基于开源 ath10k 驱动进行功能设计与实现。首先需要修改 ath10k 相关控制流程，对应图 6 中的 Soft-MAC drivers，网卡型号为高通 QCA9984。控制指定终端需要标识符，理论上可以使用 MAC、IP 和关联 ID (AID, associate ID) 作为标识。其中 AID 是无线路由器与终端关联时分配的身份标识符，与终端 MAC 一一对应，具备唯一性，且仅占用 4 个字节。考虑到嵌入式设备中内存极其珍贵，因此采用 AID 作为终端标识符。在此基础上，该主动控制方法根据模糊逻辑决策得到终端的 MCS index，将要控制的终端及其 MCS index 转化后存入用户态的指定位置，而后调用 iw 命令通过控制通道调用被修改的内核函数读取该位置内容，最后将读取后的内容转换为相应变量，并由在内核修改的函数执行，从而实现业务需求驱动的无线传输速率自适应控制。

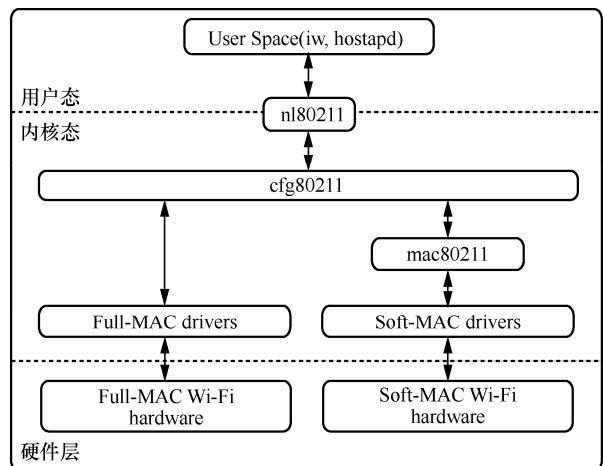


图 6 Linux 无线子系统内核分层示意

3 实验验证

本节通过真实场景实验验证了所提直播视频传输优化机制的有效性和性能。首先介绍了实验环境、实验设备等相关设置；然后在用户密集环境中，通过对比无优化机制、开启 LoL+[4]以及开启 S-Edge 这 3 种不同条件下直播视频传输业务的 QoS 指标，从而验证所提机制的有效性。

3.1 实验设置

为了构建用户密集型无线网络环境，本实验共使用了一台服务器、一台 AP、一台 PC（配备 18 块 Wi-Fi 网卡，并安装用户终端集群）。实验设备的具体配置如表 2 所示。服务器、AP 和用户终端集群之间通过以太网交换机有线连接，且用户终端集群符合用户密集环境[22]定义要求。S-Edge 部署在 AP 上，对客户端和服务器均透明。AP 使用 IEEE 802.11ac 协议工作在 5 GHz 频段上，并给 AP 分配 80 MHz 带宽信道。视频客户端使用 Selenium 自动化工具控制 dash.js 提供的直播视频播放器请求直播数据流。直播视频源由 DASH Industry Form 提供，该视频源为互联网上公开的测试视频。在评价指标方面，本文采用时延、吞吐量、丢包率 3 个常用的 QoS 指标进行量化评估。

3.2 实验方案及结果

为了验证所提机制的有效性，本文提出两套实验方案，评估在 2 种不同场景下开启 S-Edge 机制对 QoS 的影响，并与 LoL+优化机制进行比较，具体实验方案如下。

1) 直播和非直播业务并存场景

由于在用户密集环境中，对可用带宽影响最大的是文件传输类应用，因此本文设置同时接入 18 个终端，其中 17 个终端传输文件，1 个终端进行直播视频业务，分别执行无优化机制、开启 S-Edge 以及开启 LoL+，以评估 S-Edge 对直播视频业务 QoS 指标的影响，实验结果如图 7 所示。

图 7(a)描述了直播视频终端的时延随时间的变化趋势，其中对于丢失报文的时延默认设置为

1 000 ms。从图 7(a)中可以看出，当无优化机制时，多个终端设备竞争带宽，导致直播视频终端时延较高，产生大量丢包，抖动现象非常明显，直播视频业务无法正常运行。LoL+虽然在一定程度上降低了时延，但抖动程度更为剧烈，网络稳定性较差。而当开启 S-Edge 后，抖动现象显著降低，时延也在直播业务可接收范围内，直播视频终端用户体验较好。图 7(b)描述了直播视频终端在 3 种不同条件下的丢包率。从图 7(b)中可以看出，当无优化机制时，丢包率达到了 60%；当开启 LoL+时，丢包率有所降低；当开启 S-Edge 后，可以发现丢包率得到显著下降，仅为 1%。图 7(c)描述了 3 种不同机制对直播视频和非直播业务平均吞吐量的影响。从图 7(c)中可以看出，当无优化机制或开启 LoL+时，非直播视频终端均获得了较高的带宽资源；当开启 S-Edge 后，非直播视频终端的整体吞吐量得到了抑制，直播视频终端的吞吐量有小幅提升，且保持较平稳的趋势，使 AP 不会处于过载状态，各个业务终端都能正常运行。总体来说，当开启 S-Edge 后，AP 处于较为健康的负载状态，直播视频终端的 QoS 指标得到了明显的提升。

2) 多直播业务并存场景

在该方案中，本文设置同时接入 18 个终端，其中，14 个终端传输文件，4 个终端进行直播视频业务。本文分别评估 S-Edge 对两类业务的影响，并与 LoL+进行比较。图 8 展示了在无优化机制、开启 S-Edge 以及开启 LoL+3 种不同条件下，直播与非直播视频业务终端的平均时延、平均丢包率和平均吞吐量。从图 8 中可以看出，当有多个直播终端和非直播终端一起竞争时，开启 LoL+对非直播业务的带宽在一定程度上进行了抑制，改善了直播业务的平均时延和平均丢包率，但仍然存在大量丢包的情况，抖动现象明显。而开启 S-Edge 后，虽然直播业务的平均吞吐量小幅下降，但在平均时延和平均丢包率方面都表现出明显的优势。

表 2

用户密集型无线网络环境的具体配置

型号和配置	操作系统	功能
一台 x86 PC	Ubuntu 20.04 LTS 64 位	部署在公网的文件下载服务器
一台 x86 PC Intel 3168/8265/9260ac 无线网卡共 18 块	Ubuntu 18.04 LTS 64 位	作为基于多射频接口的用户终端集群，用以构建用户密集环境
一台 WPQ864 开发板（配备 QCA9984 网卡）	OpenWrt 19.07.8	作为 AP，建立和管理 Wi-Fi 热点

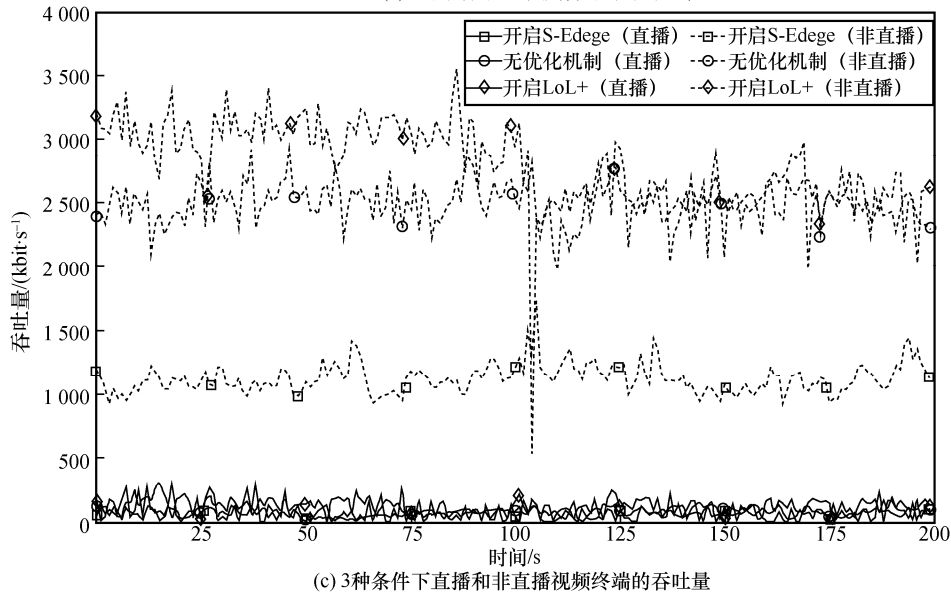
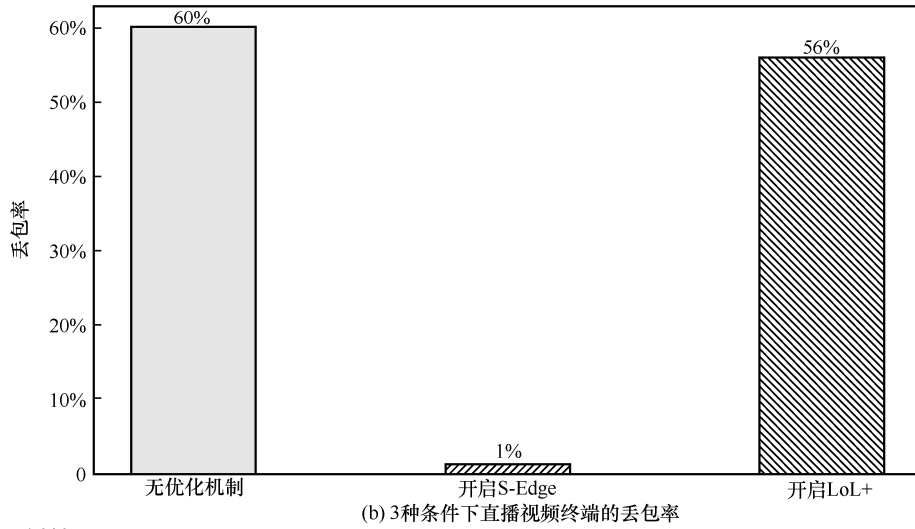
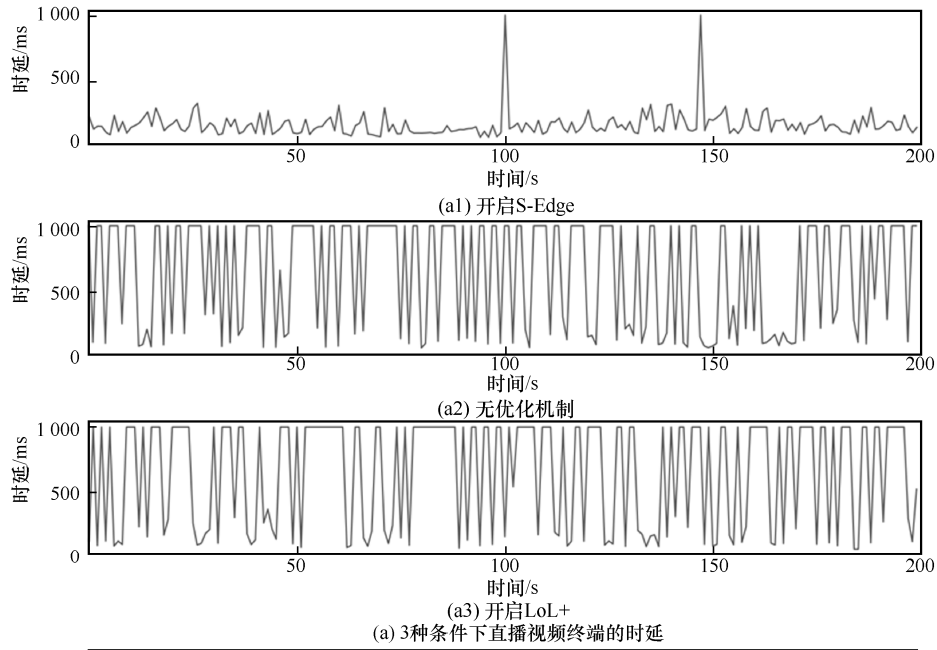


图 7 在直播和非直播视频业务并存场景中，无优化机制、开启 S-Edge 以及开启 LoL+对直播视频业务 QoS 指标的影响

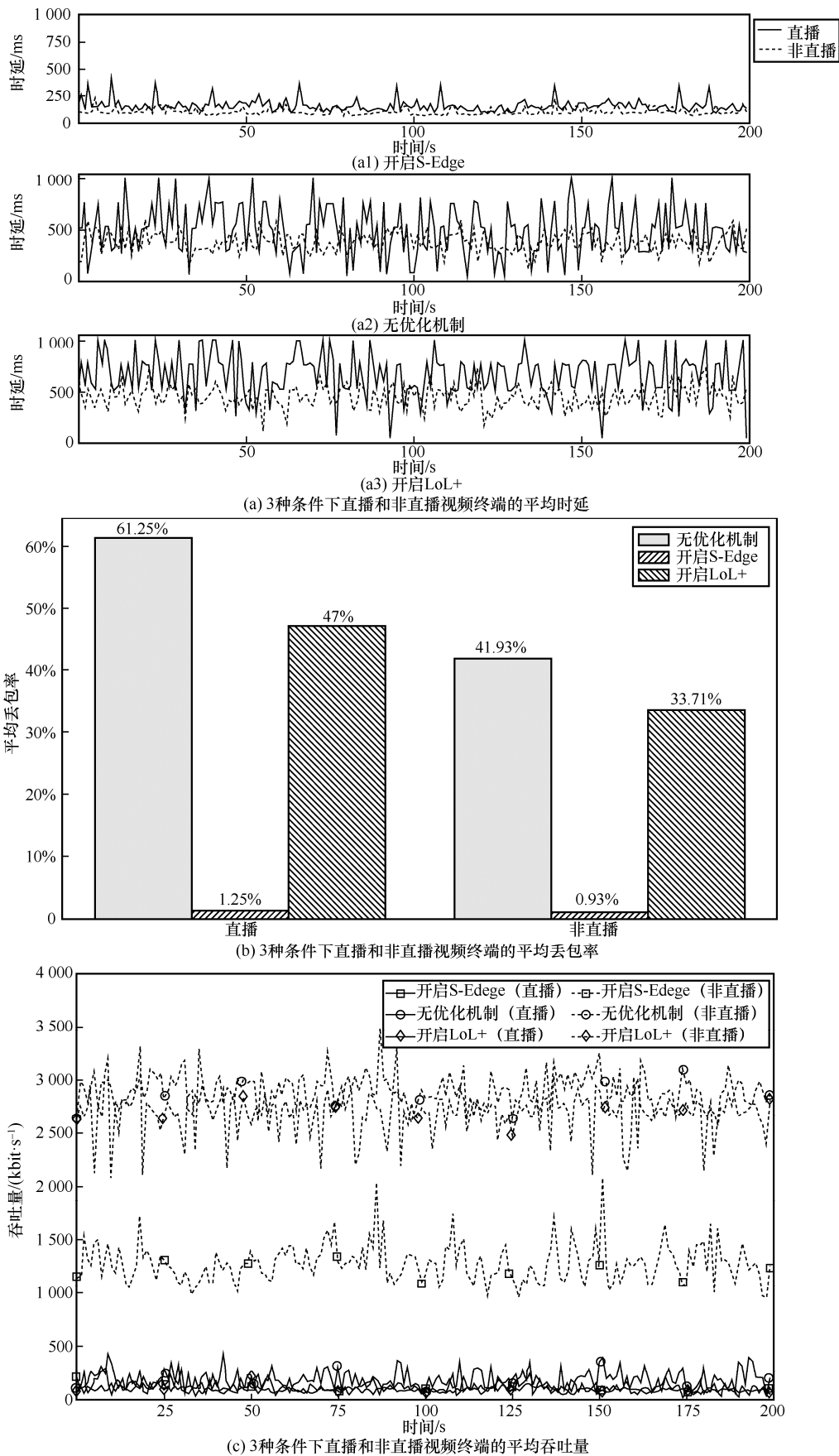


图 8 在多直播视频业务并存场景中，无优化机制、开启 S-Edge 以及开启 LoL+对直播和非直播视频业务 QoS 指标的影响

4 结束语

本文对用户密集无线网络环境下直播视频优化传输机制进行研究。针对终端间竞争带宽造成直播业务 QoS 波动明显的问题,提出了一种基于 AP 边缘智能的直播视频传输优化机制 S-Edge,并通过实验验证该机制可以显著降低平均时延和丢包率,能够满足用户密集环境下直播视频传输业务的 QoS 需求。然而,由于模糊逻辑理论需要事先定义隶属度函数和模糊规则,依赖于专家经验,缺乏可学习性。因此,在下一步工作中,将考虑引入轻量级神经网络算法来提升智能性。此外,研究基于 eBPF 的直播流量测量和调度技术,可进一步优化直播业务的传输性能。

参考文献:

- [1] HSIEH P C, LIU X, HOU I H. Fresher content or smoother playback: a Brownian-approximation framework for scheduling real-time wireless video streams[C]//Proceedings of the Twenty-First International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing. New York: ACM Press, 2020: 41-50.
- [2] ZHANG H H, ZHOU A F, LU J M, et al. OnRL: improving mobile video telephony via online reinforcement learning[C]//Proceedings of the 26th Annual International Conference on Mobile Computing and Networking. New York: ACM Press, 2020: 1-14.
- [3] ZHANG C, LIU J C, WANG Z, et al. Look ahead at the first-mile in livecast with crowdsourced highlight prediction[C]//Proceedings of IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2020: 1143-1152.
- [4] BENTALEB A, N AKCAY M, LIM M, et al. Catching the moment with LoL- in twitch-like low-latency live streaming platforms[J]. IEEE Transactions on Multimedia, 2021, 24: 2300-2314.
- [5] MU P K, ZHENG J K, LUAN T H, et al. AMIS: edge computing based adaptive mobile video streaming[C]//Proceedings of IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2021: 1-10.
- [6] WANG F X, ZHANG C, WANG F, et al. Intelligent edge-assisted crowdcast with deep reinforcement learning for personalized QoE[C]//Proceedings of IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2019: 910-918.
- [7] LUO Z X, WANG Z L, HU M, et al. LiveSR: enabling universal HD live video streaming with crowdsourced online learning[J]. IEEE Transactions on Multimedia, 2023, 25: 2788-2798.
- [8] ALI R, SHAHIN N, ZIKRIA Y B, et al. Deep reinforcement learning paradigm for performance optimization of channel observation-based MAC protocols in dense WLANs[J]. IEEE Access, 2018, 7: 3500-3511.
- [9] 胡雨涵. 视频直播传输算法的优化与系统实现[D]. 北京: 北京邮电大学, 2022.
HU Y H. Optimization and system implementation of live video transmission algorithm[D]. Beijing: Beijing University of Posts and Telecommunications, 2022.
- [10] GUTTERMAN C, FRIDMAN B, GILLILAND T, et al. Stallion: video adaptation algorithm for low-latency video streaming[C]//Proceedings of the 11th ACM Multimedia Systems Conference. New York: ACM Press, 2020: 327-332.
- [11] 乔春雨. 面向真实环境自适应视频流的用户体验优化研究[D]. 北

京: 清华大学, 2021.

- QIAO C Y. Research on user experience optimization of adaptive video streaming for real environment[D]. Beijing: Tsinghua University, 2021.
- [12] BHATTACHARYYA R, BURA A, RENGARAJAN D, et al. QFlow: a reinforcement learning approach to high QoE video streaming over wireless networks[C]//Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM Press, 2019: 251-260.
- [13] BABU M M, REDDY P C, SAM R P. A novel cross-layer based priority aware scheduling scheme for QoE guaranteed video transmission over wireless networks[J]. Multimedia Tools and Applications, 2022, 81(20): 28129-28164.
- [14] MAO H Z, NETRAVALI R, ALIZADEH M. Neural adaptive video streaming with pensieve[C]//Proceedings of the Conference of the ACM Special Interest Group on Data Communication. New York: ACM Press, 2017: 197-210.
- [15] BALAN D G, POTORAC D A. Linux HTB queuing discipline implementations[C]//Proceedings of First International Conference on Networked Digital Technologies. Piscataway: IEEE Press, 2009: 122-126.
- [16] OTT H, MILLER K, WOLISZ A. Simulation framework for HTTP-based adaptive streaming applications[C]//Proceedings of the Workshop on NS-3. New York: ACM Press, 2017: 95-102.
- [17] RACA D, SANI Y, SREENAN C J, et al. DASHbed: a testbed framework for large scale empirical evaluation of real-time DASH in wireless scenarios[C]//Proceedings of the 10th ACM Multimedia Systems Conference. New York: ACM Press, 2019: 285-290.
- [18] YI G, YANG D, BENTALEB A, et al. The ACM multimedia 2019 live video streaming grand challenge[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM Press, 2019: 2622-2626.
- [19] CAPDEHOURAT G, ÁLVAREZ G, ÁLVAREZ M, et al. High density emulation platform for Wi-Fi performance testing[J]. Ad Hoc Networks, 2018, 70: 1-13.
- [20] ARIF T Y. Evaluation of the minstrel-HT rate adaptation algorithm in IEEE 802.11n WLANs[J]. International Journal of Simulation: Systems, Science & Technology, 2017, 18(1): 1-7.
- [21] MENDEL J M. Fuzzy logic systems for engineering: a tutorial[J]. Proceedings of the IEEE, 1995, 83(3): 345-377.
- [22] YIN X Q, JINDAL A, SEKAR V, et al. A control-theoretic approach for dynamic adaptive video streaming over HTTP[C]//Proceedings of the ACM Conference on Special Interest Group on Data Communication. New York: ACM Press, 2015: 325-338.

[作者简介]



顾晓丹 (1987-), 女, 江苏常州人, 博士, 东南大学讲师, 主要研究方向为移动互联网、网络安全、匿名通信等。

吴文甲 (1983-), 男, 江苏盐城人, 博士, 东南大学副教授, 主要研究方向为移动物联网、智能无线网络等。

凌振 (1982-), 男, 江苏宜兴人, 博士, 东南大学教授, 主要研究方向为移动物联网、网络安全与隐私、可信计算等。