5G 16K 虚拟现实视频传输关键技术

张行功，郭宗明
（北京大学王选计算机研究所，北京 100871）

摘要：随着5G网络的商用普及，高质量的虚拟现实视频服务逐渐成为可能。为了达到传统高清视频的质量，需要传输16K/24K分辨率的虚拟现实视频，这对网络、终端和服务器都带来巨大的带宽、计算压力。针对这些挑战，本文介绍了虚拟现实视频传输的质量评价和面临的挑战，并详细介绍了两种自适应传输关键技术：（1）多层FOV传输；（2）多用户协作传输。重点解决自适应传输中的卡顿、黑场、多用户竞争等问题，提高虚拟现实视频传输的用户体验质量(Quality-of-Experience)。

关键词：视点自适应传输；虚拟现实视频；强化学习
中图分类号：O422　　**文献标识码：A**

5G 16K Virtual Reality Video Streaming

Xinggong Zhang , Zongming Guo

(Wangxuan Institute of Computer Technology, Peking University, Beijing 100871)

Abstract: With the deployment of commercial 5G network, Virtual Realty Video (VR Video) are coming to our life. To achieve the equal quality as HDTV, VR video needs 16K/24K resolutions which introduce significant overhead on network, server and terminal. This paper presents the quality metrics and challenges of viewport-adaptive VR video streaming. Two viewport-adaptive streaming methods are proposed: 1, Multi-layer FoV-adaptive streaming. 2, Multi-user Cooperative Streaming with reinforcement learning. They address the quality impairment of stalling, black hole and mutli-user competition and improve the quality-of-experience of VR video streaming.

Key words： viewport-adaptive streaming, virtual-reality video, reinforcement-learning

1 引言

随着5G商用时代的来临，虚拟现实技术这一最值得期待的5G应用场景得到了迅猛发展。在2019年的建国70周年阅兵式上，央视频、北京联通及华为公司就合作推出了“5G+国庆VR直播”业务；在2020年的春节联欢晚会上，中国中央广播电视台也首次启用了5G+8K/4K+VR模式拍摄并制作鼠年春晚。与此同时，国内外主流移动VR设备也在蓬勃发展。Oculus、HTC、创维、Pico、大朋都相继推出了VR一体机等移动设备。其中，创维、Pico

科技部重点研发专项：”媒体融合架构与编码传送”，　项目号：2019YFB1802701
作者简介：张行功（1973-）/郭宗明(1966-)，男（汉族），北京大学王选计算机研究所，zhangxg@pku.edu.cn / guozongming@pku.edu.cn

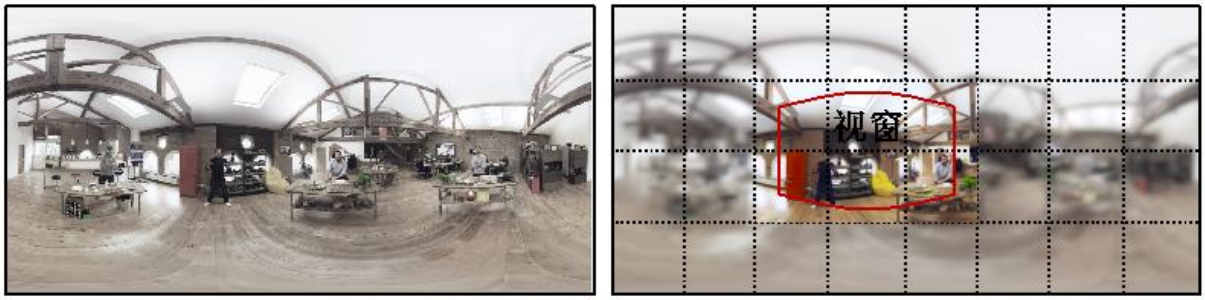

图 1 视点自适应分块传输模式：原始全景视频（左），视点区域传输（右）

的屏幕分辨率都达到了4K，视场角都超过了100°。华为、Pareal及3Glasses也都相继提供了可以与智能手机结合的轻量式VR分体眼镜，进一步提升了用户佩戴VR设备的舒适度、流畅度和持久度[1]。

即使在5G网络中传输高清的虚拟现实（VR）视频也绝非易事。相比于普通视频，VR视频需要对360°空间的内容进行采集、拼接和编码，需要更大视角和更高分辨率。以持任意的用户观看视角，以提供极致的沉浸式体验。但是，由于人眼角度及VR设备的限制，用户在每一时刻最多可观察110°左右的视窗（FoV，Field-of-View），要想在用户FoV内提供4K分辨率，则整个VR视频的分辨率至少需要达到8K甚至16K。此外，为提升用户观看虚拟现实视频的沉浸感，VR视频的帧率需要达到120帧/秒。以现有的H.265视频编码标准为例，90帧的4K超高清视频需要50-200Mbps的传输速率，120帧的8K超高清视频则至少需要200-800Mbps带宽的支持，而120帧的8KVR视频带宽需求甚至需要达到1Gbps[2]。现有的4G和WiFi网络环境完全无法支持高清全景视频的流畅播放，即使是用户体验速率在100Mbps至1Gbps的5G网络[3]也只是刚好满足要求，对于16KVR视频仍然有些无能为力。并且，随着VR视频观看人数的提升，势必也要对骨干网的带宽产生巨大的消耗。

除带宽外，VR视频对于VR观看设备的CPU、GPU效率也提出了严苛的要求。为避免用户观看VR视频时产生的眩晕感，VR视频帧的渲染延迟需要在10-20ms以内[4][5]。而降低延迟的关键就在于芯片的处理能力。为此，不少VR设备提供商如Oculus、HTC等都对用户的PC设备提出了要求。比如Oculus就明确指出用户PC设备的显卡需要支持NVIDIA GTX 1060/AMD Radeon RX 480，CPU需要支持Intel i5-4590，内存则至少需要达到8GB[6]。此外NVIDIA、AMD、Qualcomm、ARM、瑞芯微等知名公司也相继推出了自己的开发工具、渲染引擎、CPU等等。

为了降低VR视频传输和观看过程中的带宽消耗及CPU占用，自适应传输技术、视点传输(FoV-dependent)应运而生。其通过对视频进行分块压缩，根据FoV传输视点区域视频[8-10]。如图1所示，原始的高清全景视频在空间上被划分为多个视频分块进行独立编码。对于不同的网络带宽状况，用户决策每个视频块的传输码率，并选择用户视窗覆盖的分块（如图中橙色区域）。视窗范围内的区域以较高分辨率或高码率进行传输，而其他区域则以较低码率进行传输，从而保证用户可以持续观看到高清画面，网络传输总码率也可以得到降低。

本文重点介绍VR视频的视点传输面临的挑战，并详细介绍了两种自适应传输关键技术：（1）多层FOV传输；（2）多用户VR传输。针对视点自适应传输中的卡顿、黑场、多用户竞争等问题，提出了以用户质量体验(Quality-of-Experience)为目标的传输方法。

2 VR 传输质量评价标准

2.1 VR视频等效质量

由于人的视野范围在90~120度，相同分辨率的VR视频清晰度仅为传统电视的1/8左右。例如：一个4K VR视频等效于传统240P电视机的清晰度[7]，画质难于达到满意的用户质量体验。VR视频分辨率与传统电视清晰度的换算关系如下图所示：

表1 VR与电视等效分辨率

VR 视频分辨率	视点分辨率 (FoV=90度)	等效电视质量	码率

4K(3840*1920)	960*960	240P	16M
8K(7680*3840)	1920*1920	480P	64M
16K(15360*7680)	3840*3840	高清 1280*720	300M
24K(23040*11520)	5760*5760	超高清 4K	3G

为了提供满意的画质体验，VR视频至少要达到16K甚至更高分辨率。采用目前的H.264/265视频编码器，所需码率达到400M~600M，远远超出现有网络的负载能力。因此，VR视频面临编码、传输等巨大挑战。

2.2 自适应传输的用户质量体验

视点区域的自适应传输是解决上述挑战的关键技术，但它存在卡顿、视点预测错误、时延等问题，影响用户质量体验。

视点自适应传输方式需要对用户视点进行预测，以便预下载用户视窗内的高清视频内容，保证用户的播放流畅度。然而，视点预测准确度会随着预测时间的增加而不断降低。例如当预测窗口在 1s 以内时，视点预测准确度可以达到 90%以上。然而当预测窗口延长至 2s 时，预测准确度则会锐减至 70%。另一方面，为保证用户的播放流畅度，VR 设备的播放缓冲区通常需要达到 10 秒甚至更长。因此，VR 视频播放流畅度和视点预测准确度存在不可调和的矛盾，一旦预测错误，用户将面临低质量的视频内容或播放卡顿，严重损害了用户体验。

3 多层 FoV 传输

多层FoV是一种将VR视频分块、分层传输的机制。一个视频块被编码为多个质量的视频层，动态决策视频层的传输，避免视频卡顿和视点错误导致的用户体验质量QoE下降。

它主要包括三个部分：

（1）多层FoV传输，以在播放流畅度和播放质量之间实现动态平衡；

（2）动态分块视频码率决策，实现长期播放质量优化；

（3）李雅普诺夫优化理论，保证在线优化算法的近似最优；

3.1 方法介绍

为保证VR视频播放的清晰度和流畅度，原始视频切分为多个视频块(Tile)，并且每个视频块被编码为多个质量的视频层。如下图所示，当视点在A区域时，该区域的显示高质量层，其他区域显示低质量层；如果视点移动到B，相应区域的视频块被更新为高质量层。基于此，该方法可以保证用户在转动头部时质量不下降，也同时能保证观看视频的流畅性。

该方法采用了李亚普洛夫模型求解最优决策。假设用户的决策时间用k表示，并假设此时用户的缓冲区内存放了L个完整的视频内容，如下图所示，标号为l_k至$l_k + L_k - 1$，那么此时，用户就可以选择是继续向后预下载新的视频$l_k + L_k$或是更新缓冲区内现有的视频l_k至$l_k + L_k - 1$。当缓冲区长度较小时，客户端会更偏向于预取新的视频，从而保证播放的平稳性；而当缓冲区较长时，客户端则会更偏向于更新已缓存的视频，从而提升播放的质量。

假设每一次下载都会对用户带来I_k的期望质量提升，那么最大化用户的长期观看体验就可以表示为下式，其中τ_k表示每个视频的下载时长，X表示分块视频决策。

$$arg\max_{X}\frac{\sum_{k=1}^{K} I_k}{\sum_{k=1}^{K} \tau_k}$$

假设每个决策时刻的缓冲区长度为Q_k，并且用$X_{k,l_k+L_k} = 1$表示用户选择预取新的视频，$X_{k,l_k+L_k} = 0$表示更新已有视频，并且每个视频分片的长度为T，那么缓冲区在相邻两个决策点内的变化可以用下式表示：

$$Q_{k+1} = [Q_k - \tau_k]^+ + X_{k,l_k+L_k} \cdot T$$

根据李雅普诺夫漂移加罚理论

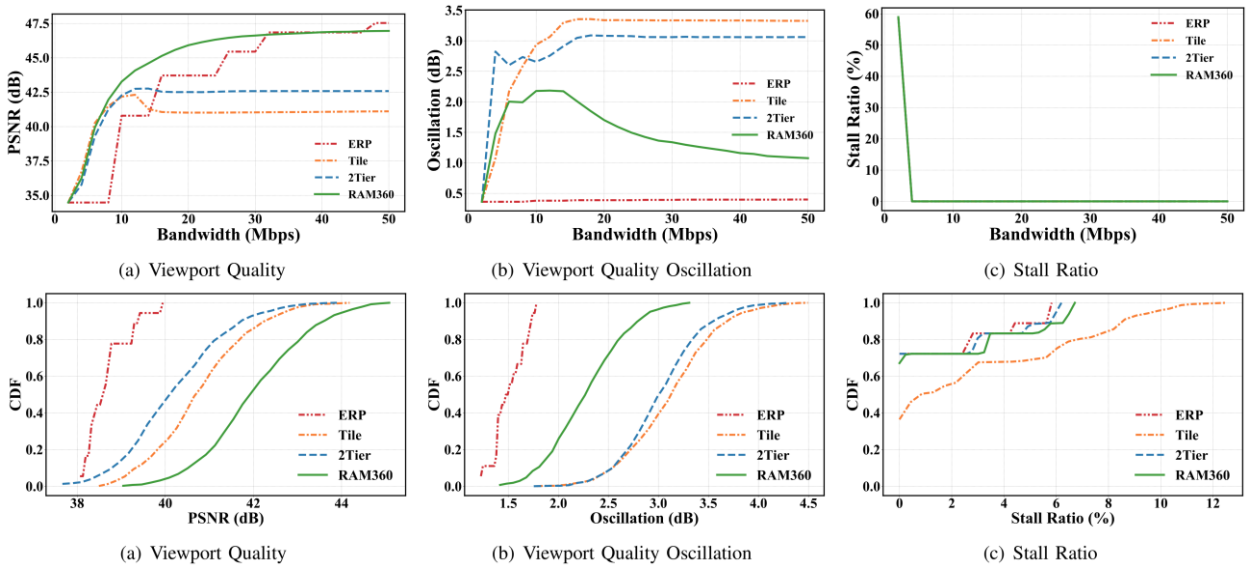

图 2 VR 视频传输 PSNR，码率抖动和卡顿比例。 (上) 不同带宽下实验结果。（下）真实网络环境

（drift-plus-penalty），为了最大化长期观看体验并且保证缓冲区的稳定性，减少播放卡顿，全局优化目标可以用一个在线优化算法进行近似替代，如下式：

$$\underset{x_k}{\operatorname{argmax}} \frac{V \cdot I_k - Q_k \cdot X_{k,l_k+L_k} \cdot T}{\tau_k}$$

其中V表示权重因子，V越大表示越重视视频质量，越小表示越重视缓冲区的稳定性。并且，利用在线决策算法进行替代，得到的解与最优解的距离是有界的，如下式：

$$\limsup_{k\to\infty} E\{I\} \geq E\{I^*\} - \frac{\Lambda + T^2}{2V}$$

3.2 实验效果

本文对比了三种传统的自适应传输方法，分别是：（1）ERP：传统的不分块的视频传输策略[6]；（2）Tile：传统的分块传输策略[10]；（3）2Tier：基于分块的双层传输方案，即先将缓冲区填满再更新视窗区域的方法[11-14]。

如图 2（上）所示，在不同带宽下，多层 FoV 传输方法都可实现最优的播放质量（以 PSNR 计），并且视频前后时间的抖动也远低于其他基于分块的视频。

在真实网络条件及真实用户视点下，对方法进行了验证，见图 2（下）。多层 FoV 方法可以实现最佳的观看质量与分块视频下最低的视频质量抖动。相比于不分块的 ERP，该方法可以提供最优的用户质量体验。上述实验充分证明，多层 FoV 传输方法可以有效适应带宽变化、视点变化等不同环境，提升 VR 视频传输质量。

4 多用户虚拟现实视频传输

为了满足VR视频的实时性、高带宽需求，VR视频已成为网络边缘的主要服务之一。当大规模用户同时观看VR视频时，由于不同用户所处的网络环境不同，视点观看区域不同，会对骨干网的带宽以及内容分发网络的缓存空间产生巨大的占用，很可能会导致网络的拥塞。

针对上述问题，边缘节点的多用户协作式传输必不可少。通过预测多用户的视点，预存多用户共同感兴趣的视频数据，从而提高边缘节点的缓存命中率，减少骨干网的视频流量和边缘缓存存储开销。

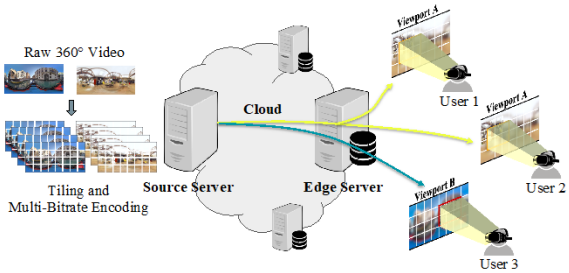

图3 多用户虚拟现实视频传输架构

传统的协作式多用户传输机制多采用中心化的自适应传输策略，即边缘节点需要承担多用户优化所带来的高昂的计算成本。为此，本文介绍一种近分布式的基于多智能体强化学习的协作传输机制，它可以：

（1）支持边缘节点协助下的大规模多用户合作，以在播放体验和服务成本之间实现动态平衡；

（2）支持在线近分布式视频码率决策，以减少边缘节点的计算压力；

（3）利用多智能体强化学习策略，以保证决策结果的近似最优性；

实验结果表明，该方法可以在小规模提升视频质量的同时减少骨干网络上40%-60%的带宽消耗及缓存占用。

4.1 多智能体协作传输机制

如下图所示，当多个用户通过相同的边缘节点接入骨干网时，每个用户都对应着不同的带宽B^k以及视窗V^k，此时网络的总带宽等于多个用户请求视频分块的并集码率之和。如果不同用户间的视窗区域可以重合，并且能够进行合理的协作，那么此时骨干网上的带宽消耗及缓存消耗就会大大降低。

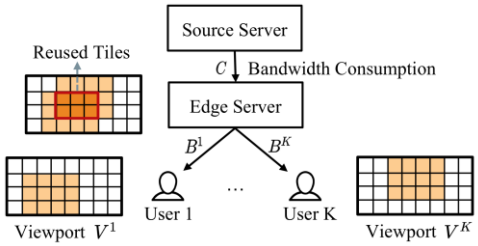

图4 多用户协作传输

假设每个用户的视频体验质量为U^k，网络总带宽占用为C，那么系统的总体优化目标就可以表示为下式，其中X表示分块视频码率决策空间：

$$\underset{X}{\operatorname{argmax}} \sum_{k=1}^{K} U^k - \eta \cdot C$$

为了实现分布式的多用户协作传输，多智能体强化学习可寻找出对用户及整个系统最优的决策。在多智能体强化学习理论中，整个交互过程可以用一个六维元组G表示，即$G = < K, S, A, R, P, \gamma >$。其中，$K$表示用户数目，$S$表示状态空间，$A$表示动作空间，$R$表示状态动作奖励，$P$表示状态转移概率，$\gamma$表示折扣因子。具体交互过程可以表述为：在每个交互时刻，用户都会同时根据自身的状态s_t^k产生不同的行为a_t^k，进而得到不同的奖励r_t^k，与此同时每个用户的状态也会根据状态转移概率p_t^k切换到新的状态s_{t+1}^k，准备进行下一次交互。在一次次交互过程中，强化学习模型最终可以学习到一个从状态到动作的最佳映射关系，也称策略π，最终实现优化目标。其中，折扣因子γ表示用户对于长期累积回报的关注程度，γ越大表示用户越关心当前动作a_t^k可以得到的长期回报，γ越小表示用户越关心当前动作a_t^k可以得到的短期回报。

为了实现分布式协作，从而降低边缘节点的计算复杂度，需要设计一种边缘节点辅助的大规模用户通信机制，如下图所示。每个客户端都将根据自己所处的网络状况和视点运动轨迹预测出下一时刻的视窗V_t^k及带宽B_t^k，并将其与当前时刻还没下载完的视频数据量d_t打包为局部状态信息$s_{l,t}^k$发送给边缘节点，而边缘节点则会将所有用户的局部状态信息进行整合（在本文中采取了取平均这一方法），并将整合后的全局状态信息$s_{g,t}$再发送给每个客户端。由此，客户端即可获得决策所需的全部状态信息s_t^k ($s_t^k = \{s_{l,t}^k, s_{g,t}\}$)，并可据此产生码率决策$a_t^k$，即视频请求$X_t$。当边缘节点收到全部用户的请求时即会产生对应的奖励信号r_t^k，并发送给每个客户端，

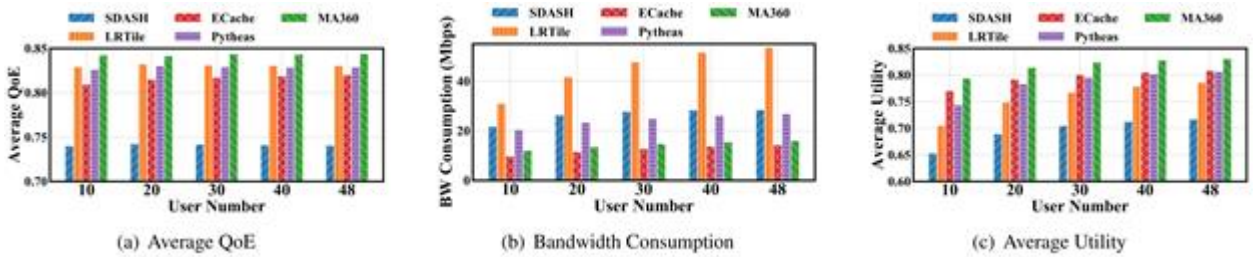

图 5 多智能体协作传输质量

从而进入下一轮迭代优化。

4.2 实验结果

对比了四种传统传输方式，分别是：（1）SDASH：即标准 DASH 传输，不将视频分割成块[15]；（2）LRTile：即利用线性回归（Linear Regression，LR）进行视点预测并传输[8]；（3）ECache：即利用边缘缓存辅助进行中心化码率分配并传输[16]；（4）Pytheas：即利用单智能体强化学习进行码率决策并传输[17]。实验结果如下：

如图 5 所示，多智能体协作机制可以在 LSTM 网络的辅助下提升视点预测的准确度，并且可以在多智能体强化学习算法的辅助下大幅减少传输所需的带宽消耗。相比于非协作式传输 LRTile，该方法可以减少 68%左右的带宽，而相比传统的 DASH 和单智能体强化学习方法 Pytheas，该方法也可以减少 41%-45%左右的带宽。

5 总结

5G 网络为 VR 视频的高质量传输提供了可能。为了满足高用户体验和清晰度，16K VR 超高清视频传输是必不可少的。它存在带宽占用高、CPU 占用率高、用户体验质量差等挑战。针对以上问题，本文介绍了多层 FoV 传输和多智能体边缘协作传输等方法，用于解决 VR 视频传输的用户体验质量优化问题及多用户视频传输的成本/质量均衡优化问题。为了进一步实现更高分辨率和质量的 24K VR 等超高清视频的传输，边缘协作、强化学习和视点传输等关键技术都将成为下一步的研究焦点。

参考文献

[1] https://consumer.huawei.com/cn/wearables/vr-glass/

[2] 陈娜,胡倩倩,曹三省.面向智能融媒体的 VR 超高清应用创新发展[J].传媒,2020(06):11-15.

[3] 徐晨霞,张洪忠.5G 条件下 VR 产业发展的突破预期[J].教育传媒研究,2020(01):30-33.

[4] Simsek M,Aijaz A,Dohler M,et al.5G-enabled tactile internet[J].IEEE Journal on Selected Areas in Communications,2016,34(3):460-473

[5] Mangiante S,Klas G,Navon A,et al.VR is on the edge:How to deliver 360videos in mobile networks[C]////Proc of the Workshop on Virtual Reality and Augmented Reality Network.New York:ACM,2017:30-35

[6] https://www.oculus.com/rift-s/

[7] 华为 VR 白皮书，https://www-file.huawei.com//media/

[8] Qian F, Ji L, Han B, et al. Optimizing 360 video delivery over cellular networks[C]//Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges. 2016: 1-6

[9] Lan Xie, Zhimin Xu, Yixuan Ban, Xinggong Zhang*, Zongming Guo, “360ProbDASH: Improving QoE of 360 Video Streaming Using Tile-based HTTP Adaptive Streaming”, Proc. of ACM Multimedia 2017, October 23–27, 2017, Mountain View, CA, USA.

[10] Yixuan Ban, Lan Xie, Zhimin Xu, Xinggong Zhang*, Zongming Guo, “CUB360: Exploiting Cross-Users Behaviors for Viewport Prediction in 360 Video Adaptive Streaming”, Proc. of IEEE International Conference on Multimedia and Expo (ICME), July 23-27, 2018. San Diego, USA.

[11] F. Duanmu, E. Kurdoglu, S. A. Hosseini, Y. Liu, and Y. Wang, “Prioritized buffer control in two-tier 360 video streaming,” in Proceedings of the Workshop on Virtual Reality and Augmented Reality Network. ACM, 2017, pp. 13–18.

[12] F. Duanmu, E. Kurdoglu, Y. Liu, and Y. Wang, “View direction and bandwidth adaptive 360 degree video streaming using a two-tier system,” in 2017 IEEE

International Symposium on Circuits and Systems (ISCAS). IEEE, 2017, pp. 1–4.

[13] L. Sun, F. Duanmu, Y. Liu, Y. Wang, Y. Ye, H. Shi, and D. Dai, "Multi-path multi-tier 360-degree video streaming in 5g networks," in Proceedings of the 9th ACM Multimedia Systems Conference. ACM, 2018, pp. 162–173.

[14] L. Sun, F. Duanmu, Y. Liu, Y. Wang et al., "A two-tier system for ondemand streaming of 360 degree video over dynamic networks," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 43–57, 2019.

[15] T. Stockhammer, "Dynamic adaptive streaming over http: standards and design principles," ACM MMSys, pp. 133–144, 2011.

[16] C. Li, L. Toni, J. Zou, et al., "Qoe-driven mobile edge caching placement for adaptive video streaming," IEEE TMM, vol. 20, no. 4, pp. 965–984, 2018.

[17] J. Jiang, S. Sun, V. Sekar, et al., "Pytheas: Enabling data driven quality of experience optimization using group-based exploration-exploitation," NSDI, pp. 393–406, 2017.