

移动网络下视频终端初始缓冲队列长度研究

房敏^{1,3}, 程光^{2,3}

(1. 东南大学计算机科学与工程学院, 江苏 南京 211189;

2. 东南大学网络空间安全学院, 江苏 南京 211189

3. 东南大学计算机网络和信息集成教育部重点实验室, 江苏 南京 211189)

摘要: 移动网络下的视频业务发展迅猛, 针对有限的移动网络带宽, 合理的规划网络资源, 提供优质的移动视频体验, 需要移动网络服务提供商对用户的视频体验质量做出准确评估。视频终端的初始缓冲队列长度是评估视频体验质量的关键参数, 对于移动网络服务提供商而言, 精确测量主流视频平台在移动网络下的队列长度值至关重要。然而, 为了保护用户隐私和网络安全, 越来越多的视频平台采用加密协议传输数据, 使得传统的 DPI 分析方法失效。本文提出一种移动网络下针对视频终端初始缓冲队列长度的研究方法, 以主流视频平台 YouTube 为例。首先, 通过关联加密视频流与播放状态的对应关系, 建立流量行为模型。然后, 构建加密视频指纹库, 支撑视频流的解析。最终, 借助模型和指纹库, 采集分析加密视频流, 测量计算初始缓冲队列长度值。实验结果表明该方法的结果能够精确到视频帧, 完全满足准确评估视频体验质量的需求。

关键词: 视频流量; 视频评估 QoE 参数; 初始缓冲队列; 加密流量分析; YouTube

Research on Video Initial Buffer Size in Mobile Network

FANG Min^{1,3}, CHENG Guang^{2,3}

(1. School of Computer Science and Engineering, Southeast University, Nanjing 211189, China;

2. School of Cyber Science & Engineering, Southeast University, Nanjing 211189, China

3. Key Laboratory of Computer Network and Information Integration of Ministry of Education of China, Southeast University, Nanjing, 211189, China)

Abstract: The video service under the mobile network is developing rapidly. For a limited mobile network bandwidth, reasonable planning of network and providing a high-quality mobile video experience, the mobile ISP need to accurately evaluate the video service quality of users. The initial buffer queue length of the video terminal is a key parameter for evaluating the quality of the video service. For the mobile ISP, it is important to accurately measure the queue length value of the mainstream video platform under the mobile network. However, in order to protect user privacy and network security, more and more video platforms use encryption protocols to transmit data, which makes traditional DPI and other analysis methods ineffective. This paper proposes a research method for the initial buffer queue length of video terminals in mobile networks, taking the representative video platform YouTube as an example. First, a traffic behavior model is established by associating the correspondence between the encrypted video stream and the playback state. Then, build an encrypted video fingerprint library to support the parsing of the encrypted video stream. Finally, with the model and fingerprint library, the collected and encrypted video data is collected and analyzed, and the estimated initial buffer queue length value is measured. Experiments show that the results of this method can be accurate to video frames, meeting the need to accurately assess the quality of video services.

Key words: Video traffic; Video evaluation QoE parameters; Initial buffer queue; Encrypted traffic analysis; YouTube

1 引言

近年来, 随着网络基础设施的不断建设以及用户数量的不断增加, 互联网业务和应用飞速发展, 尤其是视频点播服务成为主流, 使得网络中的视频

数据规模急剧膨胀^[1]。2016 年全球视频流量占有所有互联网流量的 73%, 到 2021 年这一值预计为 82%^[2]。YouTube 拥有超过 10 亿用户, 占互联网总人数近三分之一, 其中超过一半的 YouTube 观看来自移动设备^[3]。可预见的是, 对于移动网络服务提供商

基金项目: 国家重点研发计划课题 (No. 2017YFB081703), 国家自然科学基金 (No. 61602114), 中央高校基本科研业务费专项资金资助, 江苏省研究生科研与实践创新计划项目 (NO. SJCX17_0021)

Foundation Items: National Key R&D Program of China under Grant No. 2017YFB0801703, the National Natural Science Foundation of China under Grant No. 61602114, Supported by “the Fundamental Research Funds for the Central Universities,” and Postgraduate Research & Practice Innovation Program of Jiangsu Province NO. SJCX17_0021

而言,亟需解决在有限的移动网络带宽上承载大量视频数据的挑战。除了加大网络基础设施的投入外,如何提高网络的利用率,使得硬件资源得到充分利用,就需要了解用户的视频体验质量(QoE, Quality of Experience),通过准确的监测和评估 QoE,动态调整网络资源和相关参数,提高移动网络的综合利用率。

已有的关于视频 QoE 的研究表明,其决定因素是观看视频时产生的等待时间,即当视频缓冲队列为空时,终端开始缓冲数据,用户需要等待,主要发生在开始播放时的初始缓冲,和播放过程中的卡顿缓冲。分析产生初始缓冲的原因,是由于视频终端应用都会为缓冲区设置一个初始缓冲的队列长度值,只有当已缓冲的视频数据(一般为缓冲区内视频帧的个数)大于此长度,才会触发视频的播放行为。因此得到此长度的准确值,便可应用于移动网络服务提供商评估视频用户的 QoE 值。

视频服务商为了保护用户隐私和网络安全,YouTube 等视频服务商相继采用加密协议传输视频数据。加密流量的出现,导致基于 DPI(Deep Packet Inspection)的传统方法已无法测量初始缓冲队列长度值,急需针对加密视频流的 DFI(Deep Flow Inspection)等方法的解决方案。同时,现有的针对初始缓冲队列长度值的研究,往往只给出一个精确到秒的估算值,无法满足准确评估 QoE 的标准。本文提出了一种在移动网络下,准确测量视频终端缓冲队列长度值的研究方法,选取 YouTube 平台,在建立视频流与播放状态关联的流量行为模型,和构建加密视频指纹库的基础上,采集分析恰当限速下的加密视频数据,计算得出精确的初始缓冲队列长度值。实验结果表明该方法的结果能够精确到视频帧,完全满足准确评估视频体验质量的需求。

本文第 2 节综述视频 QoE 评估以及 YouTube 初始缓冲队列长度的相关研究。第 3 节具体描述了初始缓冲队列长度的研究方法。第 4 节给出了实验描述和实验结果,并对结果做了相应的分析。第 5 节总结全文并展望了未来工作。

2 相关研究

视频服务的飞速发展,推动了视频数据传输机制和视频 QoE 评估的相关研究。Hoßfeld 等^[4]提出,由于 YouTube 传输层采用了 TCP 的可靠传输,通常不会观察到视频图像的失真,QoE 主要与观看视

频时的等待时间有关,用户的等待时间可以分为初始等待,即视频开始播放之前的等待时间;以及卡顿等待,即视频播放过程中的等待时间。进而对 HTTP 承载的视频流的研究表明^{[5],[6]},QoE 的关键影响因素是初始时延和卡顿时延。为了量化用户的网络视频观看体验,华为公司联合牛津大学建立了一个用户视频观看体验的量化模型 vMOS^[7]。该模型模仿语音业务利用 MOS 量化语音质量^[8]的思想,根据在线视频初始缓冲时延(sStartDelay)、视频源质量(sQuality)、视频卡顿因素(sStalling)等客观因素来量化用户观看视频的观看体验。通过已有研究表明,初始缓冲队列长度都是 QoE 评估的重要参数。

Wamser 等^[9]指出,YouTube 终端下载的视频数据存储在内存或系统临时文件中,其作为视频播放的缓冲区,只有当缓冲了一定播放时长的视频数据后,上层播放器才会正式开始播放视频,但是 Wamser 的研究中并没有给出这一时长的具体值。Ameigeiras^[10]和 Schatz^[11]研究给出初始缓冲数据的可播放时长分别为 2.2 秒和 1.9 秒,但是这两个结果是 YouTube 在应用层采用渐进式视频下载技术^[12](HPD, HTTP Progressive Download)时研究得到的,而当前 YouTube 在应用层已弃用 HPD,改为 DASH 视频传输技术(DASH, Dynamic Adaptive Streaming over HTTP),DASH 与 HPD 在传输机制和流量特征上有很大区别。同时,视频数据的基本单位为帧,初始缓冲区实际排列的是解码后的原始视频帧队列,因此能够精确描述初始缓冲队列长度的单位应为视频帧的个数,而非可播放几秒钟的时长。本文即针对初始缓冲队列长度展开研究,精确测量队列长度值,研究结果精确到帧。

3 初始缓冲队列长度的研究方法

本文通过分析移动网络下的 YouTube 加密视频数据,精确测量 YouTube 在 Android 终端的初始缓冲队列长度值。由于原生的 YouTube APP 的限制,首先通过 YouTube Android API 开发基于 YouTube APP 之上的应用,以建立视频流与播放状态的关联关系,构建流量行为模型。加密的 YouTube 视频数据无法解构出视频帧等必要信息,为保证测量结果能够精确到帧,需构建加密视频的指纹库,用于支撑加密视频数据的解析。通过恰当的限速,采集 YouTube 的加密视频流,最终计算得出初始缓冲队列长度的精确值。

3.1 建立视频流与播放状态关联的流量行为模型

初始缓冲队列长度对应的是视频播放的初始缓冲阶段，而 YouTube 原生 APP 产生的流量无法与播放状态的建立对应关系，不能用于初始缓冲队列长度的分析研究。解决方案是借助 YouTube 官方提供 YouTube Android API 开发库，开发基于原生 APP 之上的 Android 应用，应用可以播放 YouTube 视频，同时开发库提供 onVideoStarted、onBuffering、onPlaying、onVideoEnded 四个回调函数，可以在终端获取准确的视频播放状态。

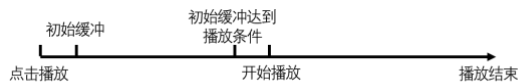


图 1 YouTube 视频的播放状态

如图 1 所示，给出了 YouTube 视频初始缓冲阶段的有关状态，表 1 列出了这些状态与 YouTube API 回调函数的对应关系。因此，两个 onBuffering 回调函数之间下载的视频数据即为初始缓冲队列长度对应的视频数据。

表 1 YouTube 播放状态与 API 回调函数关系

播放状态	对应 YouTube API 的回调函数
点击播放	onVideoStarted
初始缓冲开始	onBuffering
初始缓冲达到播放条件	onBuffering
开始播放	onPlaying
播放结束	onVideoEnded

在 Android 终端可以得到准确的播放状态，下一步需要将播放状态和加密视频流建立关联。

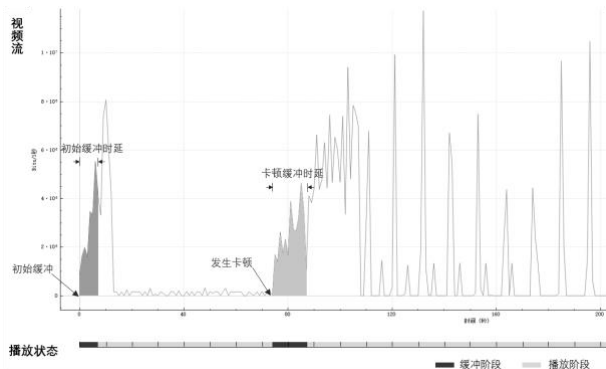


图 2 建立视频流与播放状态关联的流量行为模型

解决方案是在回调函数处主动构造状态标识报文，则视频流中就混入了播放状态的“时间点”，如图 2 所示，通过此对应关系可以在加密的视频流

中剥离出初始缓冲阶段的数据。至此，完成了视频流与播放状态的关联，流量行为模型得以构建。

3.2 构建加密视频指纹库

YouTube 视频自 2016 年开始，已全面采用加密协议传输视频数据，传统的针对 YouTube 视频流的 DPI 分析方法均已失效。为达到初始缓冲队列长度测量结果精确到帧的目的，需要构建包括视频帧偏移信息库在内的一系列加密视频指纹库。

目前已有针对 YouTube 加密视频指纹库构建方法，是对热点视频，在测试终端进行视频点播，同时在网络接入点采集视频流的报文数据。对视频请求后服务器响应的一簇密集的响应报文负载进行累加，认一簇密集的响应报文对应一个视频片段，由此构建视频的指纹库。现有的方法存在三个主要问题：

(1) DASH 机制中音、视频分离在两条流中，现有方法忽视了数据量较小的音频片段也被累加造成的数据量误差。

(2) 网络传输总会存在丢包情况，服务器和客户经过协商会重传丢失的数据包或者放弃重传，现有方法无法从加密数据中识别出丢包，因此无法考虑丢包造成对数据长度累加的误差。

(3) 网络信道条件是不断动态变化的，YouTube 采用的是自适应的传输机制，因此每次传输时音频流和视频流并行传输的特征，是随着网络状态的改变而存在差异的。但是现有的方法是对加密数据的传输进行统计，实际上获得的只是某次的传输指纹，网络环境变化后，同样的视频可能会有不同的传输指纹。

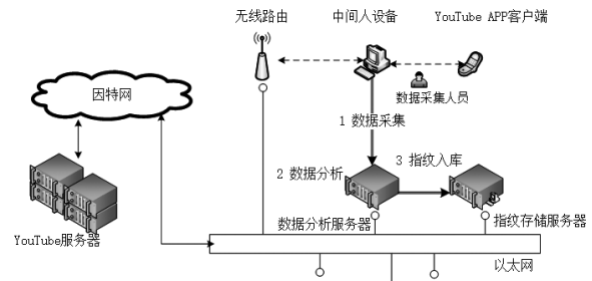


图 3 构建加密视频指纹库

针对上述问题，提出的构建加密视频指纹库方法如图 3 所示，在中间采集设备上使用中间人方法，获得目标视频的明文数据，解析采集的数据文件，获得视频的基本信息指纹，并基于 YouTube 视频分发机制和 DASH 数据传输机制进行数据分析，进一

步获得与传输状态相关的视频传输指纹，将这些指纹存入数据库中，表 2 给出了指纹库中包含的各类信息库和简要说明。

表 2 加密视频指纹库的组成

信息库名称	说明
视频基本信息库	包含视频 ID、帧率、时长、视频所有可用分辨率等信息
分片信息库	包含音频和视频分片的格式、编码方式、各分辨率下分片大小等信息
帧偏移信息库	包含各分辨率下，分片内每一帧相对分片头的偏移信息（单位 bytes）

与现有的指纹库构建方法相比，本文提出的方法解决了存在的三个问题：

(1) 本文方法是基于中间人技术，在中间节点获得视频的明文数据，通过在实验过程中切换所有分辨率的明文数据，精确读取出视频各分辨率中视频分片、音频分片的长度信息。避免了现有方法中无法区分音频分片、视频分片的问题，也避免了现有方法中无法考虑的丢包所带来的问题。

(2) 从明文中读取到的音频分片、视频分片的长度信息码，构成了视频的基本信息指纹。基于 YouTube 视频分发机制和 DASH 数据传输机制对视频的基本信息指纹分析，计算出在不同的网络信道条件下所有可能的音频和视频并行传输可能性，获得该视频在不同的网络服务质量情况下所有可能的传输指纹。

(3) 本文方法中获得的视频指纹是从中间人设备采集的明文数据中获得，与具体的传输环境无关，因此可以应用于网络性能不断变化的真实网络，具有较好的适用性。

3.3 采集加密视频流

YouTube 在传输层之上采用了 TLS 加密协议，TLS 加密会对应用层视频数据进行分块，一般为 16K 大小一块。在终端为了解密和验证每一个 TLS 块数据，必须保证 16K 加密数据都已完整收到。

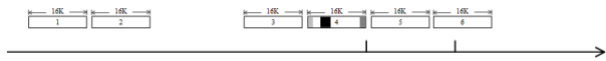


图 4 不限速下接收到 TLS 块的示意图

如图 4 所示，是不限速情况下 YouTube 视频数据到达终端的示意图，图中每一个长方块即为一个 16K 大小的 TLS 块。由于一般视频帧的大小不会大于 16K，所以初始缓冲队列长度对应的视频帧大概

率位于 TLS 块的内部（第四个长方块中的黑色小方块给出示意）。在 t_1 时间点时，从接收到的数据而言，已经达到了播放条件。但由于上层播放器需要经历一些列的解协议、解封装、解码、判断当前缓冲队列长度等一系列操作，到视频真正开始播放已经在 t_1 之后的一段时间了，即为图 4 中的 t_2 时间点。图 4 的示意图中可以看到，在 t_1 至 t_2 时间段内，又接收了第五个完整 16K 的 TLS 块。这 16K 的加密数据依旧会被当做初始缓冲的数据，在计算初始缓冲队列长度时造成结果误差。



图 5 限速下接收到 TLS 块的示意图

图 5 是限速后的 YouTube 视频数据到达终端的示意图，此时 TLS 块之间的时间间隔被拉长，示意图中 t_1' 时达到了播放条件，同样是经过一些列的操作，到 t_2' 时开始播放，至 t_2' 时间点第五个 TLS 块仍未接收完整，不会被当做初始缓冲的数据，那么队列长度就是由初始缓冲队列长度对应视频帧（第四个长方块中的黑色小方块给出示意）所在 TLS 块算出。

综合上述分析，为了保证计算的 TLS 块是初始缓冲队列长度对应帧所在的块，需要设计限速模块，排除由于终端播放器正式播放时间点滞后于真实达到播放条件的时间点，造成接收到多余 TLS 块的干扰。因此，在采集加密视频流应用中涉及了限速模块，限速的标准是既要保证视频不会因为带宽太低无法播放，也要保证排除接收到“多余” TLS 完整块的干扰。通过大量的实验证明，限速 1M 带宽，可以满足上述要求。

3.4 计算初始缓冲队列长度精确值

如 3.3 所述，通过恰当的限速，每个视频都可以定位初始缓冲队列长度对应视频帧所在的 TLS 块，如图 6 示意图，框出的真实值即为初始缓冲队列长度的精确值。

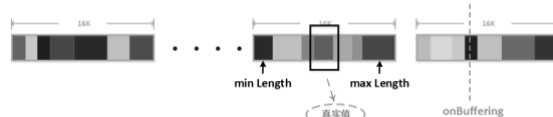


图 6 初始缓冲队列长度值所在区间

对于一个视频而言，无法直接确定真实值具体对应哪一帧，但如示意图所示，可以参照指纹库中的帧偏移信息，确定‘min Length’和‘max Length’

的具体值, 则真实值位于 $[\min \text{ Length}, \max \text{ Length}]$ 区间内部。

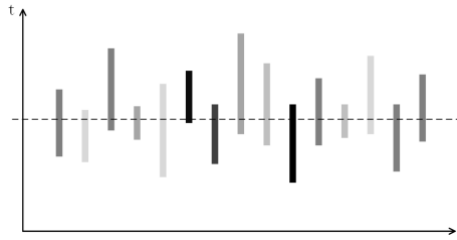


图 7 取交集计算初始缓冲队列长度值示意图

对于大量视频而言, 每一个视频都可以计算出一个初始缓冲队列长度所在的区间, 如图 7 所示, 对所有视频计算出的区间取交集, 且结果为整数, 即可得到一个尽可能精确地初始缓冲队列长度值。

4 实验分析

4.1 实验环境

实验共使用了四部测试手机, 详细参数信息见表 3。

表 3 移动设备参数信息			
设备型号	系统版本	分辨率	内存
SAMSUNG S6	Android 5.1	2560x1440	3GB
SAMSUNG Note5	Android 6.0	2560x1440	4GB
Huawei Mate8	Android 7.0	1920x1080	3GB
Huawei Mate2	Android 4.2	1280x720	2GB

全部手机均已 Root, 并在手机中安装了基于 Android YouTube API 开发的应用, 在应用中集成了 tcpdump 抓包模块, 可以在视频播放过程中自动抓取流量。

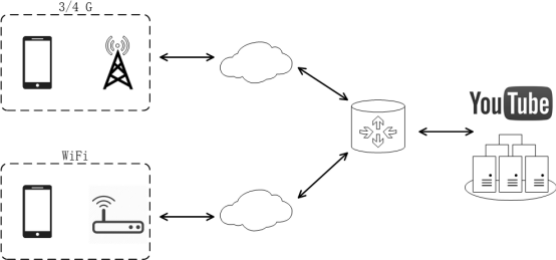


图 8 实验数据采集环境

视频流量数据采集自蜂窝网络(3/4G)和 WiFi 两种网络环境, 采集环境如图 8 所示。蜂窝网络下的数据在香港采集, WiFi 下的数据既有香港采集的数据, 也有通过代理采集的数据。

4.2 数据采集

首先, 确定实验视频集, 选取标准共三个: 一

是覆盖 0~2min、2~5min 和>5min 三种时长要求; 二是视频的开始几秒钟画面变化比较剧烈, 目的是使起始阶段视频帧的平均大小不至于太小, 导致每个 16K 的 TLS 块可能包含大量的帧, 给结果造成误差; 三是覆盖了目前 YouTube 上最普遍的三种帧率: 24FPS、25FPS 和 30FPS。最终一共确定了 61 个测试视频。

然后, 通过利用中间人工工具采集到的明文数据, 构建实验视频集的加密视频指纹库。

最后, 运行测试手机中开发的应用, 设置恰当限速环境, 自动播放和采集视频数据, 对采集到的数据分析得出结果。

实验前后共采集了四批数据: (1) 2017 年 12 月 23 日在香港采集, 蜂窝网络环境(3/4G), 约 1300 余例; (2) 2017 年 12 月 24 日在香港采集, WiFi 网络, 不限速, 约 600 余例; (3) 采集于 2018 年 3 月 8 日, WiFi 环境下, 国内代理访问, 限速 1M, 共计 560 例; (4) 采集于 2018 年 4 月 6 日, WiFi 环境下, 国内代理访问, 不限速, 共计 550 例。

4.3 实验结果分析

如表 4 所示, 给出了实验视频集中的部分结果。所有数值均为视频帧的个数。

表 4 视频集中部分视频的计算结果				
视频帧率	视频 ID	Min	Max	结果(帧数)
24 FPS	e5N_Tq1EtRQ	59	61	[60, 61]
	nx1R-eHSkfM	60	61	
	9OeZYhjyweQ	51	64	
	Mqaobr6w6_I	55	61	
25 FPS	prnQgSpV3fA	60	61	[62, 63]
	QAwL0O5nXe0	62	64	
	msUKf_cEKo0	55	66	
	OsSKFlk8oEo	59	73	
30 FPS	p-yIGS9eW-o	54	63	[75, 76]
	_9YMpuLDnwo	60	67	
	IWwhYPjyE1U	68	76	
	h_67XloYXTc	55	84	
30 FPS	JVHFMBfG-cM	75	83	[75, 76]
	iMg9ibfcBnM	74	85	
	TZSVf-yD7N8	75	76	

24FPS 的视频结果为[60, 61], 换算成时长为 [2.5, 2.542]; 25FPS 的结果为[62, 63], 换算成时长为 [2.48, 2.52]; 30FPS 的结果为[75, 76], 换算成时

长为[2.5,2.533]。因此,可以认定 YouTube 的初始缓冲队列长度换算成时长应为 2.5 秒。

进一步分析结果,初始缓冲队列长度在 YouTube 内部应该并不是一个不变的值,而应是对不同帧率的视频,有一个不同的满足播放条件所需的帧数。对于 24FPS 的视频而言,当队列长度 ≥ 60 即满足播放条件;对于 25FPS 的视频而言,当队列长度 ≥ 63 即满足播放条件;对于 30FPS 的视频而言,当队列长度 ≥ 75 即满足播放条件。

5 结束语

视频终端初始缓冲队列长度是评估用户 QoE 的重要参数,但现有关于队列长度的研究并未给出其精确值。同时,为了保护用户隐私和防止运营商干涉视频传输过程,越来越多的网络视频服务采用加密协议传输数据,使得传统 DPI 方法无法测量初始缓冲队列长度。本文提出了一种移动网络下针对视频终端初始缓冲队列长度的研究方法,以建立流量行为模型和构建加密视频指纹库为基础,采集分析加密视频流,测量计算出初始缓冲队列长度的值,结果精确到帧。下一步的工作将结合加密视频码率识别技术,利用初始缓冲队列长度值计算初始时延和卡顿时延,完善视频 QoE 评估模型。

参考文献:

- [1]. Sandvine, "Global Internet phenomena: Latin America & North America," Tech. Rep., Jun. 2016.
- [2]. Cisco, VNI, "Cisco Visual Networking Index: Forecast and Methodology 2016–2021.(2017)," Tech. Rep., 2017.
- [3]. YouTube. YouTube Press Room - Statistics. [Online]. Available: <https://www.youtube.com/yt/press/statistics.html>.
- [4]. Tobias Hoßfeld, Sebastian Egger, Raimund Schatz, Markus Fiedler, Kathrin Masuch, and Charlott Lorentzen. 2012. Initial delay vs. interruptions: Between the devil and the deep blue sea. In Proceedings of the 2012 4th International Workshop on Quality of Multimedia Experience (QoMEX'12).
- [5]. Tobias Hoßfeld, Michael Seufert, Matthias Hirth, Thomas Zinner, Phuoc Tran-Gia, and Raimund Schatz. 2011. Quantification of

YouTube QoE via crowdsourcing. In Proceedings of the 2011 IEEE International Symposium on Multimedia (ISM'11).

- [6]. Tobias Hoßfeld, Raimund Schatz, Ernst Biersack, and Louis Plissonneau. 2013a. Internet video delivery in YouTube: From traffic measurements to quality of experience. In Data Traffic Monitoring and Analysis. Springer.
- [7]. 基于 U-vMOS 视频体验标准的固定宽带网络评估建议白皮书. [EB/OL]. <http://www.c114.net/tech/167/a922663.html>, 2015.
- [8]. Nunes R D, Pereira C H, Rosa R L, et al. Real-time evaluation of speech quality in mobile communication services[C]//Consumer Electronics (ICCE), 2016 IEEE International Conference on. IEEE, 2016: 389-390.
- [9]. Florian Wamser, Pedro Casas, Michael Seufert, Christian Moldovan, Phuoc Tran-Gia, and Tobias Hoßfeld. 2016. Modeling the YouTube stack: From packets to quality of experience. Computer Networks 109, 211–224.
- [10]. P. Ameigeiras, J.J. Ramos-Munoz, J. Navarro-Ortiz, J.M. Lopez-Soler, Analysis and modelling of YouTube traffic, Eur. Trans. Telecommun. 23 (4) (2012) 360–377.
- [11]. R. Schatz, T. Hoßfeld, P. Casas, Passive youtube QoE monitoring for ISPs, in: Workshop on Future Internet and Next Generation Networks (FINGNet-2012), Palermo, Italy, 2012.
- [12]. Pastushok I, Turlikov A. Lower bound and optimal scheduling for mean user rebuffering percentage of HTTP progressive download traffic in cellular networks[C]//Problems of Redundancy in Information and Control Systems (REDUNDANCY), 2016 XV International Symposium. IEEE, 2016: 105-111.

作者简介:



房敏 (1989-), 男, 山东泰安人, 东南大学硕士生, 主要研究方向为网络测量。地址: 江苏省南京市江宁区东南大学路 2 号(211189), 手机: 13951825151, E-mail: mfang@njnet.edu.cn。



程光 (1973-), 男, 安徽黄山人, 东南大学教授、博士生导师, 主要研究方向为网络测量、网络安全和网络管理等。地址: 江苏省南京市江宁区东南大学路 2 号(211189), 手机: 13813873702, E-mail: gcheng@njnet.edu.cn。