

# 移动网络加密 YouTube 视频流 QoE 参数识别方法

潘吴斌 程光 吴桦 徐健

(东南大学计算机科学与工程学院 南京 210096)

(计算机网络和信息集成教育部重点实验室(东南大学) 南京 210096)

**摘 要** 移动视频业务应用广泛,流量占比高且持续增长.针对有限的移动网络带宽,如何合理地规划网络服务、提供优质的移动视频体验,需要客观的视频体验评估反馈网络服务提供商和视频服务运营商以改善网络利用率及传输方案.当前大多数视频服务质量评估方法都基于 DPI(Deep Packet Inspection)方法获取视频播放信息以计算视频 QoE(Quality of Experience).然而,为了保护用户隐私和网络安全,越来越多的视频采用 HTTPS 加密传输,使得传统的 DPI 方法无法获取码率和清晰度等 QoE 评估参数.因此,文中提出一种基于视频块统计特征的加密视频 QoE 参数识别方法(以代表性网络视频 YouTube 为例).首先,根据 SSL/TLS 协议握手过程中未加密部分识别 HTTPS 加密的 YouTube 流量.然后,根据视频流前若干个包的 4 种特征识别出 HLS、DASH 和 HPD 传输模式,再根据视频块统计特征建立机器学习模式识别视频块的码率和清晰度.实验结果表明该方法传输模式、码率和清晰度识别平均准确率分别达到 98%、99%和 98%,可以有效用于加密 YouTube 的 QoE 评估.

**关键词** HTTPS 视频流量;机器学习;QoE 参数识别;体验质量评估;加密 YouTube

**中图法分类号** TP393 **DOI 号** 10.11897/SP.J.1016.2018.02436

## A Method for Identifying the QoE Parameter of Encrypted YouTube Traffic in Mobile Network

PAN Wu-Bin CHENG Guang WU Hua XU Jian

(School of Computer Science and Engineering, Southeast University, Nanjing 210096)

(Key Laboratory of Computer Network and Information Integration, Ministry of Education, Nanjing 210096)

**Abstract** YouTube is one of the most popular and volume-dominant video streaming services in today's Internet. Almost 50% of the YouTube views are from mobile users, and this trend is expectedly increasing in near future. For a limited mobile network bandwidth, how to reasonably plan network services in order to provide high-quality mobile video experience? The answer requires the objective feedback of video experience for network service providers and video service providers to improve network utilization and transmission strategy. Due to the feasibility and cost effectiveness, objective video QoE (Quality of Experience) assessment is commonly used to estimate the user perception on the quality of video streaming services. Active probing can only provide instant samples but cannot accurately represent the actual network condition over the entire period of video streaming service. In contrast, passive measurement can be performed either at the client devices or in network. However, client-side measurements are more intrusive as end users are directly involved, who can provide accurate view from an individual's perspective

收稿日期:2016-03-28;在线出版日期:2017-05-24.本课题得到国家“八六三”高技术研究发展计划项目基金(2015AA015603)、江苏省未来网络创新研究院未来网络前瞻性研究项目(BY2013095-5-03)、江苏省“六大人才高峰”高层次人才项目(2011-DZ024)、中央高校基本科研业务费专项资金和江苏省普通高校研究生科研创新计划资助项目(KYLX15\_0118)资助.潘吴斌,男,1987年生,博士研究生,主要研究方向为网络安全、网络测量及流量分类. E-mail: wbpn@njnet.edu.cn.程光,男,1973年生,博士,教授,博士生导师,主要研究领域为网络安全、网络测量与行为学及未来网络安全.吴桦,女,1973年生,博士,副教授,主要研究方向为网络安全、网络测量及网络管理.徐健,男,1992年生,硕士研究生,主要研究方向为网络安全、网络测量及流量分类.

on several objective Key Performance Indicators (KPIs). Many studies regard Stallings and initial delays on the video playback as the most relevant KPIs for QoE in HTTP video streaming. In the case of adaptive streaming, quality switches is also regarded as an important factor on QoE. Although it is clear that all these KPI factors indeed have impact on QoE assessment result, there is very few QoE assessment framework proposed to systematically combine multiple KPIs to present their joint effect on QoE. Therefore, cooperated with human factors engineering lab, Huawei maps the subjective feelings to objective KPIs to establish video MOS(vMOS) evaluation framework to synthetically assess video QoE, which is composed of video source quality, initial buffering latency, and stalling ratio. However, the resulting QoE has to be estimated from traffic characteristics typically via DPI(Deep Packet Inspection), which becomes infeasible after HTTP video streaming was replaced by HTTPS. Recently, HTTPS has been adopted by major video content providers including YouTube and Netflix to provision video services to mobile users with better protection on user privacy. Adaptive streaming is also commonly used as an effective mean to enhance user QoE by dynamically adjusting bitrate suited to current network condition. The previous video quality assessment methods compute the bitrate from the video size and playback duration based on DPI or YouTube API. However, we are not able to get these parameters from the encrypted traffic when the QoE evaluation is conducted in the network. In this paper, we propose a machine learning based method for QoE parameter identification only using the traffic characteristics from the network layer. Decision tree is applied as the quick base classifier to identify the bitrate and resolution of video chunks. We have also extensively studied the impact of bitrate estimation deviation on KPI parameters used in vMOS. Experimental results show that the proposed method can be effectively used to identify the QoE parameters of encrypted YouTube video traffic, the average accuracy of transfer mode, bitrate and resolution identification reaches 98%, 99% and 98% respectively.

**Keywords** HTTPS YouTube video traffic; machine learning; QoE parameter identification; QoE assessment; encryption YouTube

## 1 引言

互联网流量监测机构 Sandvine 发布的研究报告显示,网络视频服务商 YouTube 和 Netflix 两家网站的流量已占据北美地区互联网下行流量的 50.3%。其中,约 50%的 YouTube 流量由移动设备产生,且占比将持续增长<sup>①</sup>。随着移动设备处理能力的不断增强以及移动网络技术的发展,YouTube 视频源质量也从 720p/1080p 升级到 2K,甚至 4K/8K,给网络承载带来巨大的挑战。当前有限的移动网络带宽需要承载大量的视频数据,亟需视频服务商和网络服务提供商进行协作提高网络利用率和传输效率,以保证用户的视频体验质量。针对视频体验质量问题,YouTube 对传输模式及编码方式作了较大的调整,在保证视频观看质量的同时,降低服务器的开

支、带宽开支等。另一方面,ISP 服务商对视频服务质量进行监测和评估,获得反映用户感受的视频体验质量,根据评估结果动态调整网络参数提高带宽利用率。

视频体验质量评估主要通过主观评估或客观评估来获取,由于主观评估耗时长代价高,且无法覆盖所有的网络状况及用户观看习惯,因此,越来越多视频服务商改用客观评估来克服主观评估的不足。然而,客观评估需要建立与主观感受相映射的客观视频体验质量(Quality of Experience, QoE)评估模型,再根据 QoE 评估模型的关键性能指标(Key Performance Indicator, KPI)来反映用户的视频 QoE。针对这一问题,华为通过人因工程研究得出影响用户体验质量的 KPIs,包括视频源质量、初始缓

<sup>①</sup> YouTube Press, Statistics. <http://YouTube.com/yt/press/statistics.html> 2016, 2, 15

冲时延和卡段时长,继而提出 vMOS(video Mean Opinion Score)评估模型来表示用户的视频观看体验.然而,vMOS模型中的 KPIs 参数都需要通过视频码率或清晰度来计算获取,但越来越多的视频采用加密传输,使得 KPIs 参数无法通过解析数据包获取.

视频服务商为了保护用户隐私和网络安全,防止网络服务提供商自行调整视频清晰度、插入广告等行为,YouTube 等视频服务商相继采用 HTTPS 加密传输视频数据.加密流量的引入导致基于 DPI(Deep Packet Inspection)的视频质量评估方案无法获取到码率及清晰度等视频 QoE 参数.在非加密场景下,基于 DPI 技术可以获取视频大小、可播放时长、码率等信息,准确地评估视频的视频源质量、初始缓冲时长、卡顿等 QoE 体验指标;而在加密场景下,原来的 DPI 方法失效了,亟需 DFI(Deep Flow Inspection)等适用于加密场景的技术,解决加密视频码率和清晰度获取的问题.

由于 DPI 方法无法解析加密视频流量的视频大小、可播放时长和码率等信息,导致 QoE 评估的 KPIs 参数无法获取,最终使得基于 DPI 的 QoE 评估方法失效.针对加密传输视频流量给 QoE 评估带来的影响,本文提出一种基于视频块特征的视频码率和清晰度识别方法.在深入分析自适应码流传输模式 HLS(HTTP Live Streaming)和 DASH(Dynamic Adaptive Streaming over HTTP)的基础上,根据视频块的统计特征建立模型进行识别,视频块特征从网络层提取,无需解析应用层数据获取明文特征.实验结果表明该方法具有较好的码率和清晰度识别性能,可以有效应用于加密视频 QoE 评估.本文主要的创新点包括以下几个方面:

(1) 提出基于视频块特征的码率和清晰度识别方法,由于视频块流量特征是连续型数据,采用最小描述长度方法对特征离散化;另外,样本中存在噪声数据和过拟合问题,采用悲观剪枝方法(Pessimistic Error Pruning, PEP)对决策树剪枝,优化决策树模型.

(2) 提出自适应码流传输模式的识别方法,并引入 4 种区分特征,包括 ACKNumber 分段数、SYN-ACK 到达时间间隔、SSL/TLS 协议版本、SSL/TLS 协议握手包字节数,该方法根据流的前若干个包就能识别出加密视频的传输模式.

(3) 引入 vMOS 评估模型,根据视频源质量、初始缓冲时延和卡段时长 3 个 KPIs 指标来综合评估

用户的视频 QoE,并验证基于视频块特征的 QoE 参数识别方法用于视频 vMOS 评估的可行性.

本文第 2 节综述加密视频 QoE 评估及加密流量识别的相关研究;第 3 节具体描述自适应码率视频流的传输模式及 QoE 评估模型;第 4 节描述基于视频块特征的视频 QoE 参数识别方法;第 5 节给出实验数据集、简要说明实验环境,并对识别方法进行性能分析;第 6 节总结全文并展望未来的工作.

## 2 相关研究

近年来,视频 QoE 评估被广泛研究.文献[1-2]表明卡顿和初始缓冲时延是 QoE 评估的关键性能指标,初始缓冲时延能被大部分用户接受,而卡顿对用户体验影响较大,少量的短时间卡顿都会导致 QoE 严重下降.Nam 等人<sup>[3]</sup>对移动网络下 YouTube 和 Netflix 视频流量进行了研究,结果表明网络带宽和 CPU 计算能力也是影响视频 QoE 的重要因素,文中通过不同设备及网络状况下的丢包率来评估 YouTube 视频 QoE.Seufert 等人<sup>[4-5]</sup>研究了自适应码流技术在 HTTP 视频服务中的应用,比较了不同网络状况下自适应码流与 HTTP 下载式传输模式的 QoE,自适应码流可以根据网络状况选择合适清晰度的视频块来改善视频 QoE,特别是网络性能差的情况下自适应码流传输模式明显优于 HTTP 下载式传输模式.Nam 等人<sup>[6]</sup>通过网页插件 YouSlow 检测卡顿时长来评估视频 QoE,插件收集了 40 个国家 20 000 个卡顿事件进一步分析研究卡顿对 QoE 的影响.Seufert 等人<sup>[7-8]</sup>开发客户端应用 YoMoApp,通过调用 YouTubeAPI 监控多个视频服务质量的关键性能指标,如播放状态、视频质量等级和初始缓冲等.然而,该评估方法需要安装客户端 App,且需要将性能参数返回服务器,侵犯用户隐私.Casas 等人<sup>[9]</sup>提出的评估方法 YOUQMAN 根据卡顿因素在线评估视频服务质量,该方法只需网络层统计特征,无需解析数据包.但是该方法只能处理未加密视频流,且未考虑视频源质量等其他 KPIs 指标对 QoE 评估的影响.本文着重关注基于 HTTPS 的视频流量,所处理的数据对象明显不同,且采用的 QoE 评估模型也更全面.

当前加密流量识别方法主要基于机器学习<sup>[10-11]</sup>和主机行为<sup>[12-13]</sup>.机器学习方法基于流统计特性建立识别模型,如流持续时间<sup>[14-15]</sup>,包数<sup>[16]</sup>,最小、最大、均值和方差的到达时间间隔<sup>[17]</sup>,有效载荷大

小<sup>[17-18]</sup>, 比特率<sup>[18-19]</sup>, 往返时间<sup>[19]</sup>, 包方向或服务器发送比特率<sup>[20]</sup>. 然后, 大部分特征对视频流 QoE 参数识别不适用. 例如: 视频流的数据包常以最大报文段长度 (Maximum Segment Size, MSS) 传输, 且视频流传输需要高带宽, 重传发生率高. 此外, TCP 参数 (如服务器发送比特率, 包到达时间间隔, 往返时间和数据包方向) 对于视频 QoE 参数识别都是弱相关特征. 文献[21]提出基于随机指纹的 SSL/TLS 应用识别方法, 根据 SSL/TLS 握手过程未加密部分的消息类型建立马尔科夫链识别模型对 13 种常见的 SSL 应用进行识别, 取得了较好的分类性能. 文献[22]提出一种基于信息熵的实时加密流量识别系统 Iustitia, 根据不同类型 (二进制流、文本流和加密流) 的熵值特点实现精细化识别. 不同于研究<sup>[21-22]</sup>, 由于自适应码流会根据网络状况动态切换视频清晰度, 不仅要识别出具体的视频应用 (如 YouTube, Netflix 和 Vimeo), 还要对每个视频块进行识别, 识别出每个视频块的 QoE 参数, 如码率和清晰度. 文献[23]针对 Safari 浏览器下观看 YouTube 视频提出一种清晰度识别方法, 准确率达到 97.18%, 但当前网络视频主流是基于自适应码流的传输模式, 针对固定清晰度的识别方法意义有限.

当前 YouTube 客户端的视频流量都采用 HTTPS

加密传输, 传统的 DPI 方法很难获取文件大小、可播放时长、码率和清晰度等信息, 暂未发现针对加密视频 QoE 参数的识别方法及相应的客观评估方法. 其中, 视频 QoE 评估的 KPIs 指标都要通过码率和清晰度等视频 QoE 参数来计算获取, 如何有效识别基于自适应码流传输模式的加密视频 QoE 参数对于用户视频体验评估至关重要.

### 3 自适应码流视频 QoE 评估模型

#### 3.1 自适应码流传输模式

YouTube 视频多年来一直沿用基于 HTTP 的视频流传输技术. 自适应码流技术对于同一个视频会配置几种不同码率的样本文件, 从而根据用户的网络状况和设备的 CPU 计算能力来匹配不同码率的视频样本进行传输. 视频文件被切割成很多小的分段, 然后基于 HTTP 按序传输给客户端, 这种传输方式存在浪费带宽的问题. 比如: 当客户端按序下载其中一个视频片段时用户切换了清晰度, 那么设备在切换清晰度后会重新下载受影响的片段. 而自适应码流会根据网络状况自动切换清晰度<sup>[24]</sup>, 视频片段传输如图 1 所示, 该传输模式不仅节约服务器网络带宽, 还有利于提高用户体验.

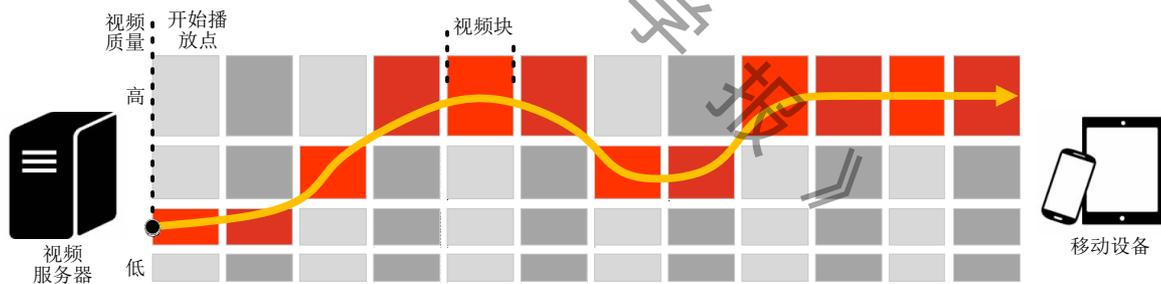


图 1 自适应码流传输模式

当前 YouTube 客户端的自适应码流传输协议有 Apple HLS 和 MPEG DASH. HLS 传输模式采用多次请求多次下载的方式, 视频流分为 video 和 audio 两种资源传输, 每个视频片段作为独立单位进行请求, 但 video 和 audio 两种资源并不是交替传输的. 视频片段大部分是根据时间等分的, 分段时长约为 5s, 也有少部分视频片段时长为 2s~10s, 且每个片段对应一个 URL, 在第一个视频分段请求前需先请求分段索引文件 m3u8. 视频传输一开始会快速传输约为 2s 的片源数据量, 有利于降低初始缓冲时延和卡顿事件. 对同一个视频而言, 不同清晰度的 video 片段数据量差别较大, 而 audio 片段数据量基

本相同, 且 video 与 audio 片段个数相同. 由于视频是分块传输的, 一个视频块数据 (video 和 audio) 由于 MTU (Maximum Transmission Unit) 限制, 会拆分为很多 TCP 数据包传输, 同一个视频片段的所有 TCP 数据包的 ACK Number 是相同的. 根据 ACK Number 可以对视频数据进行分块整合, 得到一系列数据块. 与 HLS 不同的是, DASH 传输模式的 video 和 audio 块是按顺序交替请求的, 视频片段大部分是根据时间等分的, 分段时长约为 10s. 视频媒体按照 Fragment MP4 (FMP4) 方式组织, FMP4 将大的媒体数据分块, 每个视频块可被单独解码并播放.

### 3.2 视频 QoE 评估模型

当前 YouTube 自适应码流传输模式 HLS 和 DASH 都采用 HTTPS 加密传输,使得 DPI 技术很难获取到视频大小、可播放时长、码率等信息,因此,无法得到视频源质量、初始缓冲时延和卡顿占比等 QoE 评估要素.为了提高带宽利用率和观看体验,需要对 YouTube 视频服务质量进行有效的评估,并根据评估结果调整网络参数.对于语音业务,由于主观评估方法成本高费时长,通常采用客观评估获取 MOS 分评估语音质量.类似的,可以采用客观评估方法来评估视频体验质量.为了提出准确有效的客观评估模型,华为根据人因工程借助眼动仪和生理仪测量实验者在观看移动视频过程中的感知情况,研究发现:移动视频源质量、初始缓冲时延与卡顿时长是影响用户体验的最重要外部因素,最终对上述三大关键影响因素与实验者主观感受映射实现视频体验质量评估量化,提出 vMOS 评估模型,vMOS 模型的 KPIs 包括视频源质量,初始缓冲时延和卡顿时长占比,将这些 KPIs 用分数 1~5 来反映用户的视频体验<sup>①</sup>.

实验结果表明影响视频源质量的关键因素包括:视频编码(Video Codec, VC)压缩算法,如 H.264, H.265, VP9;视频源编码等级(Coec Profile, CP),如 Base, Main, High Profile;视频清晰度(Video Resolution, VR)和码率(Video Bitrate, VB).视频源质量 sQuality<sup>①</sup>表示如下:

$$sQuality = Quality_{max} * \left(1 - \frac{1}{(1 + VB * VC * CP / VR)^2}\right) \quad (1)$$

其中,  $Quality_{max}$  代表当前清晰度视频源质量评分的最高值,对于清晰度为 4K, 2K, 1080p, 720p, 480p 和 360p 的视频,  $Quality_{max}$  值分别为 4.9, 4.8, 4.5, 4, 3.6 和 2.8.

初始缓冲时延(Initial Buffering Latency, IBL)是从点击“播放”到视频播放的等待时间,较差的网络状况会导致较长的视频缓冲时延,这种缓冲对视频服务质量所产生的影响是视频 QoE 评估的重要指标.初始缓冲时延 sLatency<sup>①</sup>表示如下:

$$sLatency = \begin{cases} 5, & IBL \leq 0.1 \\ 1, & IBL > 10 \\ 0.25 + 4.66 * e^{-IBL/5.37} \end{cases} \quad (2)$$

卡顿是指播放过程中缓冲的数据量被消耗完所产生的视频播放停止,卡顿事件对于用户体验影响很大,与卡顿时长、次数等因素有关,主要由

卡顿时长占比(Stalling Ratio, SR)决定.卡顿时长占比 sStalling<sup>①</sup>表示如下:

$$sStalling = \begin{cases} 5 - 20 * SR, & SR \leq 0.15 \\ 0, & SR > 0.45 \\ 2 - 20 * \frac{SR - 0.15}{3} \end{cases} \quad (3)$$

综合考虑视频源质量、初始缓冲时延、卡顿占比 3 个 KPIs 提出 vMOS 模型对视频 QoE 进行客观评估,vMOS<sup>①</sup>表示如下:

$$vMOS = (1 - 0.092 * (1 + 2e^{-sLatency})) * (5 - sLatency) * sQuality - 0.018 * (1 + 2e^{-sStalling}) * (5 - sStalling) \quad (4)$$

## 4 基于视频块的视频 QoE 参数识别

### 4.1 视频 QoE 参数识别系统

图 2 描述了基于视频块特征的视频 QoE 参数识别系统架构.系统主要由 4 个模块组成:HTTPS YouTube 流量识别模块,传输模式识别模块,视频 QoE 参数识别模块,以及视频 QoE 评估模块.首先,HTTPS YouTube 流量识别模块根据 ClientHello 数据包中“googlevideo”字段预先建立的视频服务器

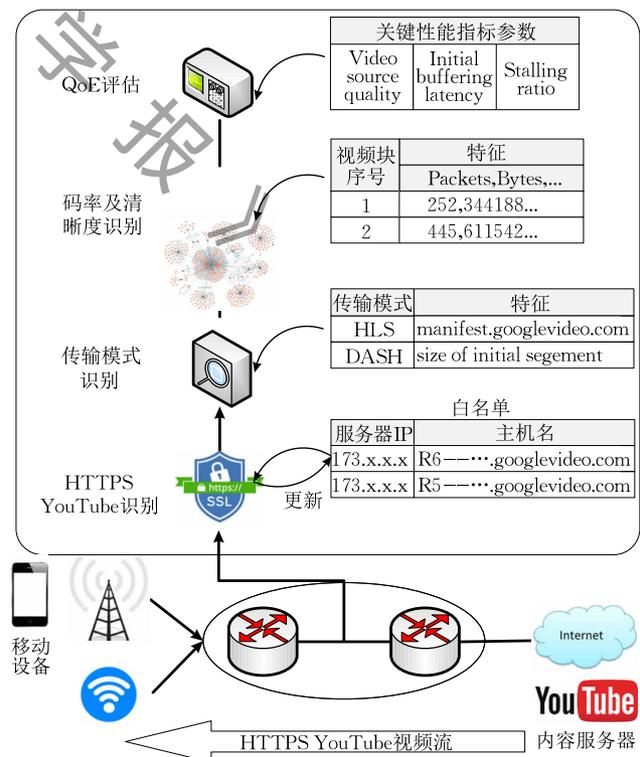


图 2 视频 QoE 参数识别及 QoE 评估系统

① Huawei. Mobile MOS. <http://www.mblab.com:9090/mobilemos/index.php?r=site/index> 2016, 2, 15

IP 白名单过滤出加密 YouTube 流量, 然后, 传输模式识别模块根据视频流前几个包的特征识别出 YouTube 流量的传输模式. 接着, 视频 QoE 参数识别模块根据获取的视频块特征分别对 HLS 和 DASH 模式的流量特征建立模型识别出视频码率和清晰度. 最后, 视频 QoE 评估模块根据视频 QoE 参数及视频流量传输参数计算 KPIs 和 vMOS 分数.

#### 4.2 HTTPS YouTube 视频流量识别

未加密 YouTube 流量通过 DPI 方法可以获得 KPIs 参数, 而加密 YouTube 流量首先需要根据加密特性从 HTTPS 流量中识别. 因此, 加密 YouTube 视频 QoE 评估的首要任务是识别出 SSL/TLS 协议下的 YouTube 流量. YouTube 视频自适应码流机制下, 视频传输前会与 YouTube 服务器交互取得媒体描述文件, 基于当前网络状态选择合适码率的视频分片, 再根据描述文件中 BaseURL 地址取得视频分片. 在非加密情况下, 根据视频分片的 IP 和用户的 IP 就可以关联视频. 然而, 在加密情况下媒体描述文件的内容是无法获取的. 为了快速关联 HTTPS YouTube 视频流量, 建立视频服务器的 IP 白名单实现快速识别, 并引入自动更新机制. 通过 DNS 响应和 SSL/TLS 握手协议中 ClientHello 可以有效获得 YouTube 视频服务器 IP. 每当读取数据报文后, 检查 Google 的 DNS 响应报文是否存在 `r*.googlevideo.com` 字段, 如果是视频服务器地址的 DNS 响应, 则查找列表中是否存在该视频服务器 IP, 如果该 IP 不存在, 就更新名单. 否则, 继续解析 SSL/TLS 握手消息中的 ClientHello 报文, 如果含有 `r*.googlevideo.com` 字段, 则看该报文的 IP 是否存在于列表中, 如果该 IP 未曾出现, 就更新名单. 最后, 通过布隆过滤器查询白名单中视频服务器 IP 与用户 IP 组成的 IP 对关联视频流.

#### 4.3 传输模式识别

当前加密 YouTube 自适应码流传输模式主要有 Apple HLS、MPEG DASH 和 HPD (HTTP Progressive Download), 不同传输模式采用不同的视频 QoE 参数识别模型, 因此, 需要在一开始就识别出传输模式. 文中基于 ACK Number 分段数、SYN-ACK 到达时间间隔、SSL/TLS 协议版本、SSL/TLS 协议握手包字节数 4 种特征采用机器学习算法进行识别, 该方法优点在于根据流的前若干个包就能识别出传输模式.

在 DASH 传输模式下, 服务器需要先向客户端发送 Initial Segment, Initial Segment 包含了视频解

码器所需的初始化信息, 然后再开始传输视频数据, 在加密数据包中的表现就是经过 SSL/TLS 握手阶段后开始传输的前  $P$  个 Application data 数据包出现  $S$  种 ACK Number, 该特征是区分视频传输模式的重要信息. 以 DASH、HLS 和 HPD 这 3 种传输模式为例, 对每类视频 100 个加密视频抽取前 3 个 Application data 数据包出现  $S$  种 ACK Number 的结果进行统计, 统计发现 DASH 前 3 个数据包的 ACK Number 种类为 2 或 3, 而 HLS 和 HPD 的种类都为 1. 该特征差异性强, 且发生在视频数据传输早期, 可以很大限度的避免重传带来的不利影响.

对比加密情况下不同传输模式的流级特征发现: 在同一时间, HPD 采用单流传输 (视频和音频没有分离), 而 DASH 和 HLS 采用多条流传输. 进一步发现 DASH 传送视频时总是两条流开始、两条流结束 (传输过程中由于网络原因导致端口更换而更换两条流中的另一条继续传输). 与 DASH 明显不同, HLS 会频繁更换流 (具体表现为变换端口) 来完成整个视频的传输. 以这 3 种传输模式为例, 随机抓取了 100 个视频前两条流的 SYN-ACK 到达时间间隔进行统计, 发现 DASH 的前两条流间隔是最短的, HLS 其次, HPD 前两条流 SYN-ACK 到达时间间隔要远远大于 HLS 和 DASH, 该特征是 HPD 与其他传输模式最明显的差异.

SSL/TLS 握手过程中包头未加密信息常被用于 SSL/TLS 应用识别. 表 1 展示了不同机型观看视频的对比结果, 可以发现同一种机型在 WEB 端或 APP 端观看 YouTube 视频时, 服务器端握手过程数据包总大小存在规律性. 同一种传输模式下数据包总大小的差别不大, 但是其不同传输模式和机型有所差异.

表 1 不同传输模式握手数据包总大小和协议版本

传输模式	协议版本	握手数据包总大小	机型
DASH-APP	TLS1.2	197/2317	HTC M7
HLS-APP	TLS1.2	207/2207	iPhone 6
HPD-APP	TLS1.0	2323/2394	华为 P7
DASH-WEB	TLS1.2	197/2317	P7-Chrome
HPD-WEB	TLS1.2	197/2134	ip6-Safari

传输模式分类过程中, 由于朴素贝叶斯算法要求属性条件独立, 然而, 实际分类过程中属性之间存在依赖关系, 使得朴素贝叶斯算法无法适用, 而贝叶斯网络不要求所有的属性都条件独立. 鉴于传输模式识别特征之间存在一定的相关性, 贝叶斯网络可适用于传输模型识别, 贝叶斯网络构建包括以下两

个步骤:

(1) 创建贝叶斯网络结构, 变量之间构成有向无环图 DAG (Directed Acyclic Graph). DAG 中的结点表示变量, 弧代表依赖关系. 采用基于评分搜索的 K2 搜索算法, 最大父节点个数设为 1, 评分方法采用最小描述长度方法 (Minimum Description Length, MDL).

(2) 训练贝叶斯网络, 构建条件概率表 (CPT). 每个变量都对应一个 CPT, 初始概率设为 0.5. CPT 表示条件分布  $P(x | Parents(Y))$ ,  $Parents$  代表双亲.

设  $X = (x_1, x_2, \dots, x_n)$  是属性  $Y_1, Y_2, \dots, Y_n$  描述的数据元组, 联合概率分布为:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | Parents(Y_i)) \quad (5)$$

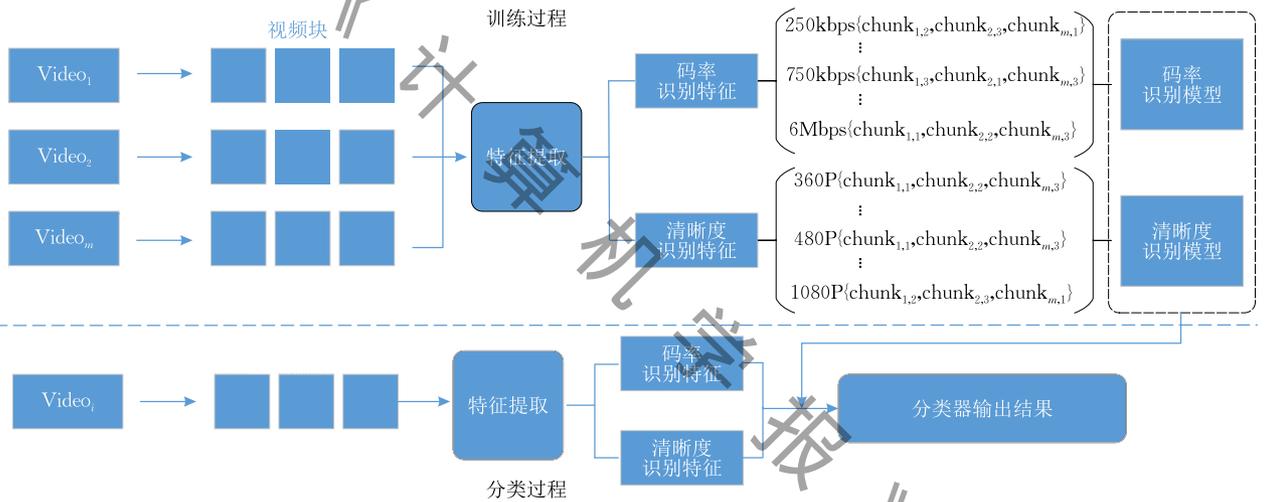


图 3 机器学习视频 QoE 参数识别流程

#### 4.4.1 YouTube 流量分段处理

DASH 和 HLS 视频传输都是将视频进行分段后使用基于 TCP 的 HTTPS 流传输的, 由于 TCP 的 MSS 所限制, 视频片段被分割成大量 1.4 KB 左右的数据包传输, 而这些数据包都是响应的同一个请求, 所以它们的 ACK Number 相同, 如图 4 所示. 根据 ACK Number 是否相同, 可以将 YouTube 加密视频流量进行分段处理.

由于视频服务器与客户端之间除了传输音视频数据还会传输目录文件以及其他交互信息, 但是这些信息数据量都远小于音视频的数据量, 所以可以使用阈值  $L$  (根据统计值默认设置为 20 KB) 来过滤非音视频数据片段. 音视频片段区分时, 音频跟视频的码率相比, 音频码率相对固定, 播放过程中保持不变, 不同清晰度的音频片段的数据量取值会集中在

其中,  $P(x_1, x_2, \dots, x_n)$  是  $X$  值的特定组合的概率, 而  $P(x_i | Parents(Y_i))$  的值对应于  $Y_i$  的 CPT 表目.

#### 4.4 视频 QoE 参数识别

文献[25-26]表明决策树算法可以有效用于流量识别, 文中采用 C4.5, RandomForest (RF) 决策树算法进行性能分析, 并与 Bayes Networks 和 Adaboost 算法进行对比. 图 3 描述了机器学习视频 QoE 参数识别流程, 包括视频块特征提取, 模型训练和样本分类. 视频块特征提取模块统计视频块网络层信息, 无需解析数据包应用层信息. 模型训练模块分别选取码率识别和清晰度识别特征子集并建立相应的识别模型. 最终, 分类模块根据建立的模型识别出视频块的码率和清晰度.

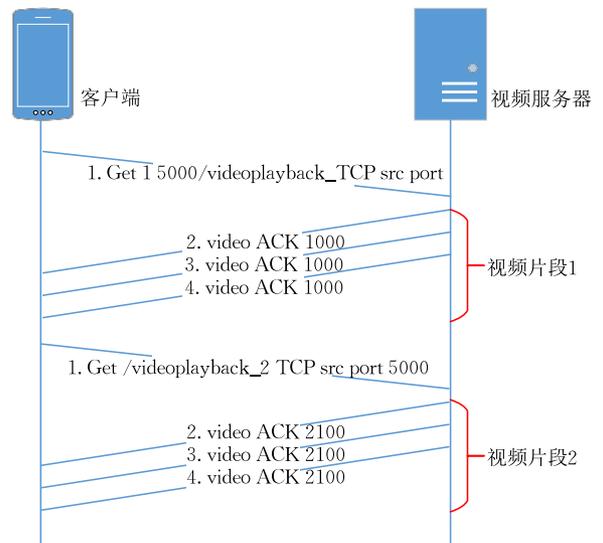


图 4 视频分段传输网络表现

一个固定的区间内。

针对异常音视频片段,在网络状况理想的情况下,一个音视频片段会在同一个 TCP 流中传输,且该片段的所有 TCP 报文的 ACK Number 相同.但是由于网络状况的不可控性,实际传输过程中会发生 TCP 流中断重传的情况,这直接导致音视频片段数据会通过不止一个 TCP 流传输.通过使用 Fiddler 中间人观察明文以及分析密文数据文件发现每一条传输加密音视频数据的 HTTPS 连接在异常中断传输时都会出现 SSLAlert 信息,出现 SSLAlert 消息的 TCP 流传输的最后一个音视频片段实际是没有传输完成而终止的.根据使用 Fiddler 对 DASH/HLS 明文消息分析得知,DASH 处理断流情况的方案是重新使用一条新的 TCP 流从上一个未传输完成的音视频片段的断点处续传数据;HLS 处理断流情况是重新使用一条新的 TCP 流将上一个未传输完成的音视频片段重新传输一次.因此,可以对带有 SSLAlert 消息的音视频片段进行拼接(DASH)或者去重处理(HLS),从而得到最终的音视频集合分别为  $H(a_1, a_2, a_3, \dots, a_k)$  和  $H(v_1, v_2, v_3, \dots, v_k)$ .

鉴于视频块或音频块由多个相同 ACK Number 的 TCP 包构成,可以根据 ACK Number 还原视频块,并统计视频块的特征,包括视频块包数、字节数、持续时间、块时间间隔、比特率、Get 请求响应时间、重传包数和源端口.本文相关研究中讨论了当前加密流量识别最常用的特征,如包长、包到达时间间隔、RTT 和数据包方向.然而,在视频流中有效载荷大小通常是 MSS,视频流在网络中的延迟差异也很大.文献[27]表明自适应码流在开始播放前会快速传输一小块视频(Initial burst),该视频块会尽快传完,减少初始缓冲时延和卡顿事件,该视频块与当前网络状况和码率具有一定的相关性.自适应码流传输中,视频块与音频块是一一对应的,鉴于音频块的码率是固定的,可以根据音频块字节数估计视频块的播放时间,防止个别播放时长不定的视频被误识别,如播放结束时或切换码率时的视频块.另外,最后一段视频块通常由数据剩余构成,且该视频流处于末尾对 QoE 评估影响不大.因此,不对该视频块进行识别.

表 2 描述了码率识别和清晰度识别的特征子集.码率识别的特征子集包括 bpackets, bbytes, burst\_bytes, audio\_bytes 和 duration. Bpackets 和 bbytes 表示视频块的包数和字节数,由于视频块的

播放时长是基本相同的,所以码率越大视频块包数和字节数也越大. Burst\_bytes 表示视频在开始播放时会快速传输一小块视频,会尽快传完,该快传视频块字节数跟码率相关性强. Audio\_bytes 表示音频块的字节数,视频块与音频块是一一对应的,因此,鉴于音频块的码率是固定的,可以估计视频块的播放时长,防止有些播放时长不定的视频被误识别,如播放结束时或切换分辨率时的视频块. Duration 表示视频块传输的持续时间,持续时间越长传输速度越小,自适应码流机制会根据传输速度选取合适的码率,以免发生卡顿事件.清晰度识别特征包括 bpackets, bbytes, duration, resolution 和 brate, resolution 表示上一视频块的清晰度,brate 表示比特率,自适应码流机制会根据比特率选取合适的清晰度,使得卡顿事件与观看体验达到平衡.该特征子集区别性强,而且特征之间也不存在冗余,可以有效的适用于视频 QoE 参数识别.

表 2 码率识别和清晰度识别的特征子集

简称	特征描述	码率	清晰度
bpackts	视频块字节数	是	是
bbytes	视频块包数	是	是
burst_bytes	初始快传字节数	是	—
audio_bytes	音频字节数	是	—
duration	视频块持续时间	是	是
resolution	上一视频块清晰度	—	是
brate	比特率(视频服务器到客户端)	—	是

#### 4.4.2 属性离散化

当连续属性多,且任一属性的取值多时,决策树复杂度将大大增加.表 2 可以看出特征都是连续型数据,连续型数据作为决策树节点,会产生很多分支,影响决策树生成和分类效率.本文采用最小描述长度方法 MDL 对数据离散化<sup>[28]</sup>.

MDL 采用描述语言长度来表示模型复杂度,目标是实现低复杂性和高准确性.描述语言越长,准确性更高;描述语言越短,模型复杂度越低.根据数学描述,MDL 模型的目标就是最小化描述语言长度  $M_{mdl}$ <sup>[29]</sup>:

$$M_{mdl} = \arg \min_{M_i \in M} \{ |L_m(M_i)| + |L_c(D|M_i)| \} \quad (6)$$

其中,  $|L_m(M_i)|$  表示模型所需的位数,  $L_m(M_i)$  表示模型  $M_i$  的描述语言,  $M_i \in M$ ;  $L_c(D|M_i)$  表示模型  $M_i$  描述对象  $D$  的语言,  $|L_c(D|M_i)|$  表示其对应描述语言所需的长度.每个对象可以看作由一个确定序列和一个随机序列构成.  $M_i$  表示确定性序列;随机序列表示确定性序列与对象的误差;确定序列可以选用自回归模型或多项式模型来表示,描述语言

长度  $|L_m(M_i)|$  为模型中参数的个数; 随机序列可以选用概率分布模型来描述, 根据香农定理, 随机过程用概率分布模型表示时,  $|L_c(D|M_i)|$  描述长度是以 2 为底对数的负数.

#### 4.4.3 决策树剪枝

如果分类样本和模型训练样本一致, 模型分类训练集样本的准确率高, 但分类新样本时准确率会下降. 因为训练样本中会存在噪声数据和孤立点, 决策树模型建立过程中会拟合这些异常样本, 使得分类模型对新样本的识别性能会下降. 因此, 有必要对决策树进行剪枝, 获得高泛化能力的分类规则. 本文采用 PEP 算法进行剪枝处理, 因为 PEP 算法在修剪过程每棵子树最多遍历一次, 效率高, 适用于大规模样本集.

PEP 方法采用连续性校正提高样本的识别可靠性. PEP 方法中如果式(7)成立, 则  $T_i$  应被剪裁:

$$e'(t) \leq e'(T_i) + S_e(e'(T_i)) \quad (7)$$

其中,  $e'(t) = e(t) + 1/2$ ,  $e'(T_i) = \sum e(i) + N_i/2$ ,  $S_e(e'(T_i)) = |e'(T_i)(n(t) - e'(T_i))/n(t)|^{1/2}$ ,  $e(t)$  代表节点  $t$  的误差,  $e(i)$  代表叶子节点  $T_i$  的误差,  $N_i$  代表子树  $T_i$  的叶子节点个数,  $n(t)$  代表训练样本数目. PEP 方法剪枝过程中最多只需访问每棵子树一次, 时间复杂性与未剪枝树的非叶子节点数目成线性关系.

## 5 实验分析

### 5.1 实验环境

移动设备流量数据采集来自蜂窝网络和 WiFi 两个网络环境, 采集环境如图 5 所示. 蜂窝网络下的数据在华为 MLAB 实验室的 4G 基站上采集, WiFi

环境下流量通过热点共享给 Android 和 iOS 终端采集, 具体设备如表 3 所示. 为了验证识别方法的鲁棒性, 文中抓取不同传输模式、不同时长、不同清晰度的视频作为测试样本, HLS 传输模式的视频通过 iPhone 客户端获取, DASH 传输模式的视频通过 Android 浏览器和客户端获取. 采集短视频、中视频和长视频 3 种类型, 短视频时长不超过 5 min, 中视频介于 5~10 min, 超过 10 min 的是长视频; 并抓取固定分辨率(360p, 480p, 720p, 1080p)和自适应码率的视频. 为了验证识别方法对于网络多样性的适应能力, 视频数据从多个国家采集, 包括韩国、巴西、中国香港和中国上海, 具体视频块样本及视频个数分布如表 4 所示, 括号中对应韩国、巴西、中国香港和中国上海采集视频的数目.

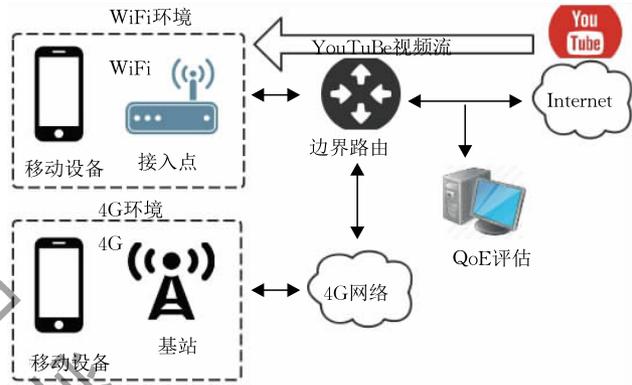


图 5 YouTube 视频质量评估框架

表 3 移动设备信息

设备	系统版本	清晰度	屏幕尺寸	内存
iPhone 6s	iOS9.1	1334×750	4.7	2G
iPhone 6	iOS8.3	1334×750	4.7	1G
iPhone 5	iOS8.3	1136×640	4.0	1G
HTC M7	Android4.4.3	1920×1080	4.7	2G
三星 S4	Android4.4.2	1920×1080	5.0	2G
小米 MI2	Android4.1	1280×720	4.3	2G

表 4 视频样本分布(视频块样本/视频个数)

传输模式	视频时长/min	清晰度				自适应码流
		360p	480p	720p	1080p	
HLS	≤5	5042/120	4217/100	3274/80	2561/60	6138/100
		(20,10,10,80)	(20,10,10,60)	(10,10,10,80)	(10,10,10,30)	(10,10,10,70)
	5<&.≤10	8490/100	6518/80	4857/60	3526/40	7916/80
		(20,10,10,60)	(20,10,10,40)	(10,10,10,30)	(10,10,10,10)	(10,10,10,50)
	>10	15371/100	12036/80	9858/60	6295/40	11324/60
		(20,10,10,60)	(20,10,10,40)	(10,10,10,30)	(10,10,10,10)	(10,10,10,30)
DASH	≤5	3714/150	3287/120	3019/100	2871/80	3502/100
		(20,10,10,110)	(20,10,10,80)	(10,10,10,70)	(10,10,10,50)	(10,10,10,70)
	5<&.≤10	5241/130	4902/100	4583/80	4225/80	5161/80
		(20,10,10,90)	(20,10,10,60)	(10,10,10,50)	(10,10,10,50)	(10,10,10,50)
	>10	8582/120	7391/90	6752/70	6382/70	7092/60
		(20,10,10,80)	(20,10,10,50)	(10,10,10,40)	(10,10,10,40)	(10,10,10,30)

由于视频数据是加密的,涉及隐私保护,当前还未有权威的公共数据集用于评估加密视频 QoE 参数识别的性能.为了使识别方法具有可比性,文中对

YouTube2016 年度 10 大视频<sup>①</sup>进行评估,每个视频采集 360p\480p\720p\1080p 以及自适应码流 5 种类型,视频具体信息如表 5 所示.

表 5 YouTube 2016 年度 10 大视频样本信息

序号	视频	时长	DASH	HLS
1	Adele Carpool Karaoke	14:52	84	167
2	PPAP(Pen Pineapple Apple Pen)	1:09	7	13
3	What's inside a Rattlesnake Rattle?	6:06	35	69
4	Nike Football Presents: The Switch ft. Cristiano Ronaldo, Harry Kane, Anthony Martial & More	5:58	34	67
5	Grace VanderWaal: 12-Year-Old Ukulele Player Gets Golden Buzzer-America's Got Talent 2016	5:25	31	61
6	Water Bottle Flip Edition   Dude Perfect	7:28	42	84
7	Channing Tatum & Beyonce's "Run The World (Girls)" vs. Jenna Dewan-Tatum's "Pony"   Lip Sync Battle	4:43	27	53
8	Donald Trump: Last Week Tonight with John Oliver (HBO)	21:54	123	246
9	The \$21,000 First Class Airplane Seat	9:05	51	102
10	Brothers Convince Little Sister of Zombie Apocalypse	3:41	21	42

## 5.2 标记方法

Fiddler<sup>②</sup> 工具通过开启本地代理进行抓包,能够解密 HTTPS 流量.首先 Fiddler 作为客户端跟服务端建立 SSL/TLS 连接,使用服务端证书处理请求和响应;然后 Fiddler 又作为服务端跟客户端建立 SSL/TLS 连接,使用 Fiddler 的证书处理请求和响应. Fiddler 解密 HTTPS 流量需要先把它生成的根证书添加到系统受信任的根证书列表中.通过解密 HTTPS 流量获取视频的明文信息,找到视频分段内容,提取分段内容中每一段的视频长度和可播放时长,计算出相应的视频码率,从而对视频块样本标记码率.通过实测发现移动设备上不同清晰度(240p,360p,480p,720p,1080p,2K)相应的平均码率为 250 kbps,450 kbps,700 kbps,1.5 Mbps,3 Mbps,6 Mbps,因此,将机器学习的样本标记设置为 250,350,450,550,700,900,1200,1500,1800,2200,2600,3000,3500,4000,5000,6000 kbps.

## 5.3 评估策略

为了验证传输模式识别、码率识别及清晰度识别方法的有效性,采用准确率 (*accuracy*) 和 *F-Measure* 综合评估识别方法的性能:

$$accuracy = \sum_{i=1}^m (TP_i) / \sum_{i=1}^m (TP_i + FN_i) \quad (8)$$

$$precision = \frac{TP_i}{TP_i + FP_i} \quad (9)$$

$$recall = \frac{TP_i}{TP_i + FN_i} \quad (10)$$

$$F\text{-Measure} = \frac{2 \times precision \times recall}{precision + recall} \quad (11)$$

评估过程中,首先对采集的视频数据进行预处理,提取视频块统计特征;然后,根据 Fiddler 的明文信息对相应的视频块标记传输模式、码率及清晰度;评估系统在 Eclipse 平台上调用 weka API 再开发实现;最后,采用十折交叉验证传输模式、码率和清晰度识别性能.

## 5.4 识别结果

### 5.4.1 传输模式识别准确率

为了验证传输模式识别方法的有效性,采用识别准确率 (*accuracy*) 和 *F-Measure* 统计传输模式识别准确性,将 4 种常用的机器学习方法 (Random-Forest, Bayes Networks, C4.5 和 AdaBoost) 进行对比,AdaBoost 采用 C4.5 作为基分类器,Random-Forest, C4.5 和 AdaBoost 采用 PEP 方法进行剪枝,DASH、HLS 和 HPD 的视频样本均为 500 个,采用十折交叉进行验证,识别准确率和 *F-Measure* 分别如表 6 和图 6 所示.

表 6 传输模式识别准确率 (单位:%)

传输模式	方法			
	Random Forest	Bayes networks	C4.5	Adaboost
DASH	97.5	99.5	97.5	97.0
HLS	96.0	96.1	97.3	95.9
HPD	99.3	99.3	98.0	98.0

从表 6 可以看出,传输模式识别准确率均高于 95.9%,因为特征中 ACK Number 分段数、SYN-ACK

① Top Trending Videos of 2016 (Global). <https://www.youtube.com/playlist?list=PLSLtz8jpJdr5qbDTyWb1bTQOe9B9s-L7trD>, 2016, 12, 7

② Telerik. Fiddler, the free web debugging proxy. <http://www.telerik.com/fiddler> 2016, 2, 15

到达时间间隔、SSL/TLS 协议版本、SSL/TLS 协议握手包字节数都是强相关特征,机器学习方法可以较好的利用相关性识别加密 YouTube 视频的传输模式.从表 6 和图 6 可以看出 BayesNetworks 的平均准确率和  $F$ -Measure 都高于其他算法,因为贝叶斯网络与决策树不同,是一种有效的概率推理模型,隐含着网络节点之间的相关性.HLS 识别准确率低于 DASH,因为 HLS 和 HPD 的 ACK Number 种类特征相似,HLS 容易被误识别为 HPD 模式.

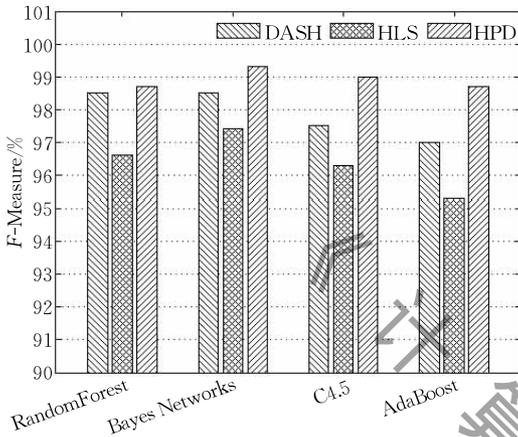


图 6 传输模式识别  $F$ -Measure

#### 5.4.2 码率识别准确率

为了验证码率识别方法的有效性,采用识别准确率对 HLS 和 DASH 传输模式的视频分别验证,由于 HPD 属于渐进式下载,与流媒体服务器每次传输 5~10 s 左右的视频数据不同,HPD 视频服务器会不停的传输数据直到视频数据下载完成,文中暂不对 HPD 传输模式进行识别.将 4 种常用的机器学习方法(RandomForest, Bayes Networks, C4.5 和 AdaBoost)进行对比,HLS 视频块样本数为 144 012 个,DASH 视频块样本数 109 624 个,采用十折交叉进行验证,结果如图 7 所示.

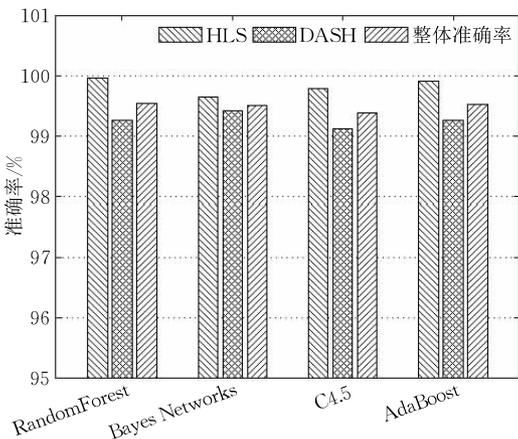


图 7 码率识别准确率

从图 7 可以看出,基于视频块特征的码率识别准确率均高于 99.1%,因为特征中视频块字节数和音频字节数与码率是强相关特征,机器学习方法可以较好的利用相关性识别加密 YouTube 视频的码率.另外,DASH 模式识别准确率低于 HLS 是由于 DASH 片段时长约为 10 s,而 HLS 的片段时长约为 5 s,片段时长越长单一视频块数据量浮动越大,容易误识别为相邻的码率类型.

为了使码率识别方法具有可比性,采用 YouTube2016 年度 10 大视频进行验证,视频块码率识别错误个数如表 7 所示.

表 7 视频块码率识别错误个数 (单位:个)

序号	DASH	HLS
1	1	0
2	0	0
3	0	1
4	0	1
5	0	1
6	1	0
7	0	0
8	1	3
9	0	0
10	0	0

从表 7 中可以看出 DASH 和 HLS 码率识别错误个数分别为 3 和 6,表明识别方法可以有效地进行码率识别.

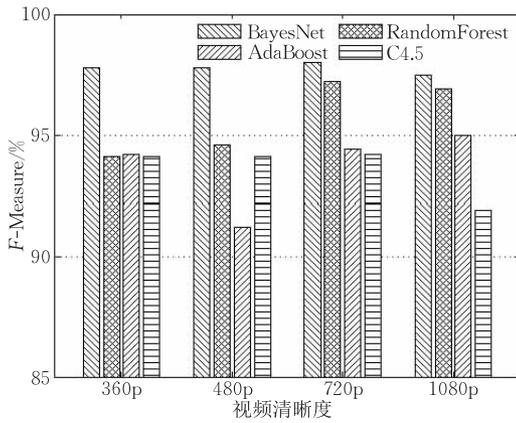
#### 5.4.3 清晰度识别准确率

表 8 描述了 4 种机器学习算法的清晰度识别准确率.识别准确率只能综合评价整个数据集的识别准确率,一个好的算法不仅要有较高的识别准确率,还应该对各个清晰度都具有较好的识别性能,特别当各个清晰度的样本分布不均匀时,在每个清晰度上的识别效果特别重要,综合评价  $F$ -Measure 可以有效地描述各类别的识别性能,结果如图 8 所示.

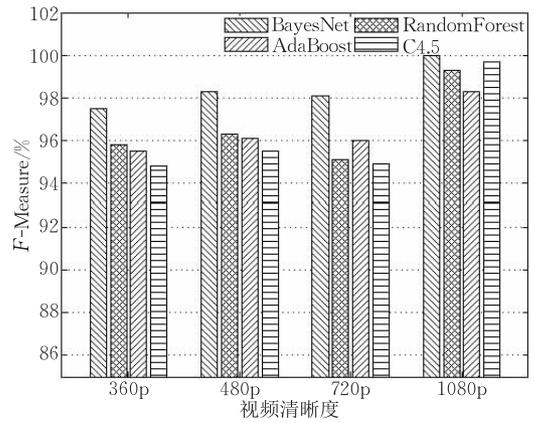
表 8 清晰度识别准确率 (单位:%)

方法	AdaBoost	C4.5	Bayes networks	Random Forest
HLS	97.80	95.78	93.70	93.70
DASH	98.18	96.09	96.09	95.49
整体准确率	97.99	95.94	94.90	94.60

从表 8 可以看出机器学习方法都取得了较好的识别准确率,因为特征中视频块的字节数和上一视频块清晰度都是强相关特征. AdaBoost 方法的准确率最高,HLS 模式的识别准确率为 97.8%,DASH 模式的识别准确率为 98.18%,因为 Adaboost 分类器对同一数据集训练不同的弱分类器,综合多个弱



(a) 传输模式HLS的F-Measure



(b) 传输模式DASH的F-Measure

图 8 清晰度识别 F-Measure

分类器获得较好的分类性能；而 RandomForest 分类器根据不同特征子集建立分类器，分类性能差的分类器影响了最终结果。从图 8 可以看出 AdaBoost 在各个清晰度都取得较好的识别效果，借助 AdaBoost 方法集成学习的优势，可以有效提高清晰度识别性能。

为了使清晰度识别方法具有可比性，采用 YouTube2016 年度 10 大视频进行验证，清晰度识别结果如表 9 所示。

表 9 视频块清晰度识别错误个数 (单位:个)

序号	DASH	HLS
1	2	3
2	0	1
3	0	2
4	1	0
5	0	2
6	0	0
7	1	3
8	3	5
9	1	2
10	1	4

从表 9 中可以看出 DASH 和 HLS 清晰度识别错误个数分别为 9 和 22，准确率分别为 98.02% 和 97.57%，表明识别方法可以有效地进行清晰度识别。

#### 5.4.4 机器学习码率识别误差

由于机器学习的码率标记是离散型数据，所以与真实码率存在一定的误差。码率误差 =  $|\text{真实码率} - \text{机器学习码率}| / \text{真实码率}$ 。为了验证机器学习对于不同时长视频带来的码率误差，选用短视频(5 min 以内)，中视频(10 min 以内)，长视频(超过 10 min，大部分为 15~30 min)进行对比分析，机器学习码率识别误差如图 9 所示。

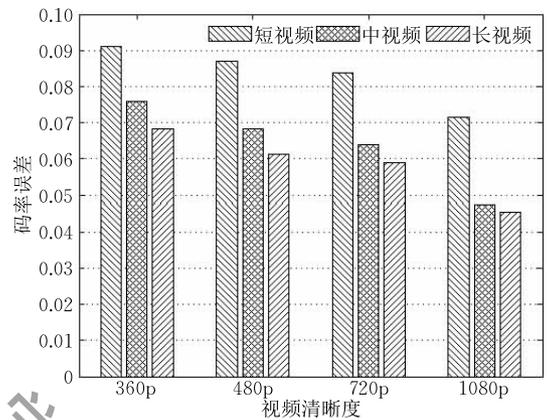
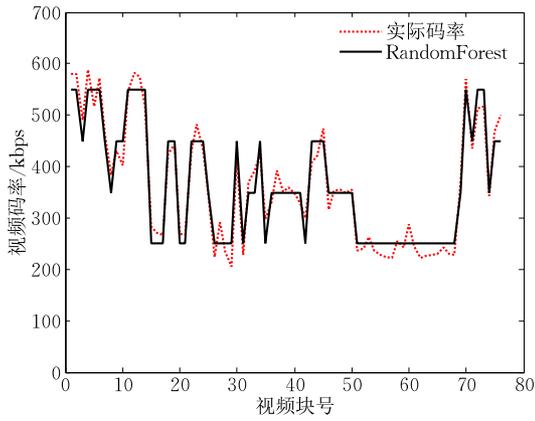


图 9 机器学习码率识别误差

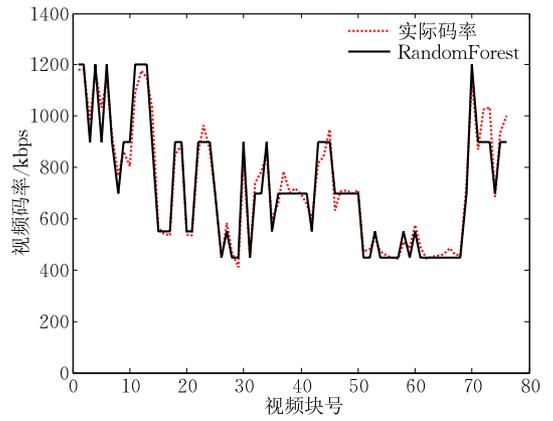
从图 9 可以看出，视频清晰度越低，误差相对越大，最大为 9.1%。因为视频本身码率较小，因此，误差相对越大。另外，中视频和长视频的误差差异随清晰度逐渐变小，当清晰度为 1080p 时最小平均误差为 4.4%，因为本身码率较大，因此误差相对较小，但误差值较大。为了具体描述实际码率与机器学习码率识别的误差，图 10 给出了视频“Cristiano Ronaldo—A Great Person”不同清晰度的码率误差。

图 10 可以看出基于机器学习的码率识别与真实码率差异较小，未出现明显差异，对不同清晰度的“Cristiano Ronaldo—A Great Person”视频码率识别均取得较好的性能，且可以长时间保持较好的识别性能，并有效的用于码率变化较大的视频，因为采用属性离散化和决策树剪枝可以构建高泛化能力的分类器。

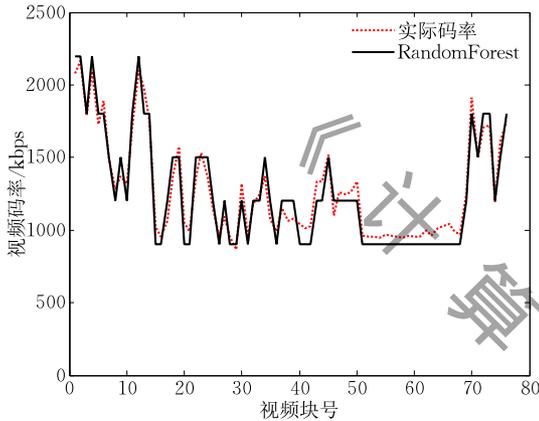
为了使码率识别方法具有可比性，采用 YouTube2016 年度 10 大视频说明码率识别误差，码率识别误差如表 10 所示。



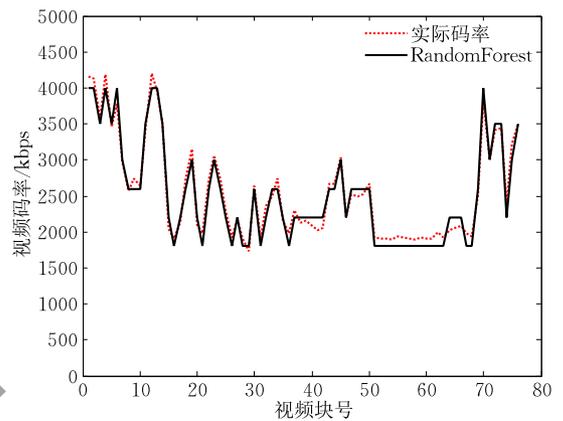
(a) 360p视频“Cristiano Ronaldo—A Great Person”  
真实与机器学习码率误差



(b) 480p视频“Cristiano Ronaldo—A Great Person”  
真实与机器学习码率误差



(c) 720p视频“Cristiano Ronaldo—A Great Person”  
真实与机器学习码率误差



(d) 1080p视频“Cristiano Ronaldo—A Great Person”  
真实与机器学习码率误差

图 10 视频“Cristiano Ronaldo—A Great Person”真实码率与机器学习码率误差

表 10 码率识别误差

序号	DASH/%	HLS/%
1	0.6	1.3
2	1.0	1.7
3	0.7	1.6
4	1.2	1.4
5	0.8	1.2
6	0.3	0.5
7	1.1	0.4
8	0.9	1.0
9	1.7	1.3
10	1.6	1.3

从表 10 可以看出码率识别方法的误差介于 0.4%~1.7%，DASH 模式的最低误差为 0.3%，HLS 模式的最低误差为 0.4%，表明该识别方法可以有效用于加密视频码率识别。

#### 5.4.5 码率误差对 KPIs 的影响

码率误差不仅影响视频源质量，还对卡顿相关和缓冲感知因素产生较大的影响，最终影响 vMOS 评估。下面分别从视频源质量、卡顿相关、缓冲感知和 vMOS 评估描述码率误差的影响。

#### (1) 码率对视频源质量的影响

视频源质量评估如式(1)所示，Qualitymax 表示视频清晰度最高达到的 vMOS 分值，VC 表示视频编码，当前 YouTube 大部分是 H.264 编码，VP 表示编码等级，有 High, Main 和 Base profile，当前 YouTube 主要是 High profile。VR 表示视频清晰度的最低码率要求，4K, 2K, 1080p, 720p, 480p, 360p 分布对应 3000 kbps, 1400 kbps, 700 kbps, 350 kbps, 200 kbps, 180 kbps。VB 代表视频码率，其余参数都是固有属性，不受码率影响，因此，视频源质量影响

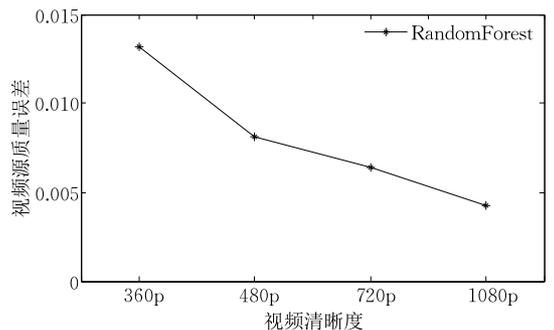


图 11 不同清晰度的视频源质量误差

只需考虑码率因素,图 11 描述了清晰度对视频源质量的影响.为了具体描述码率对视频源质量的影响,图 12 描述了视频“Cristiano Ronaldo—A Great Person”不同清晰度的码率误差对视频源质量的影响.

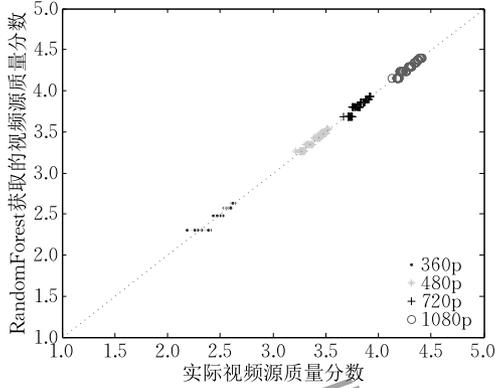


图 12 不同清晰度的视频“Cristiano Ronaldo—A Great Person”的视频源质量误差

从图 11 可以看出,视频源质量平均误差随视频清晰度增加而下降.因为视频源质量与清晰度有关,不同清晰度对应不同的 vMOS 基础值也不同.由于清晰度越低,视频源质量评估基础值越小,导致相同的视频源质量分值变化对低清晰度视频影响相对较大.

### (2) 码率对初始缓冲时长的影响

流媒体播放等待时延具体反映当用户观看 YouTube 视频时,从点击播放按钮开始,到视频开始播放的时长.初始缓冲感知评估如式(2)所示,IBL 表示初始缓冲时延,IBL ≤ 0.1, 1, 3, 5, 10 s 分别对应初始缓冲时长分值的 5, 4, 3, 2, 1. 然而,初始缓冲时延 = 码率 × 2 s / 初始缓冲平均速率,也与码率相关,统计初始缓冲速度为 2 Mbps, 4 Mbps, 8 Mbps, 16 Mbps, 32 Mbps 时对初始缓冲感知的影响,如图 13 所示.为了具体描述码率对初始缓冲时延的影响,图 14 描述了视频“Cristiano Ronaldo—A Great Person”不同清晰度的初始缓冲时延影响.

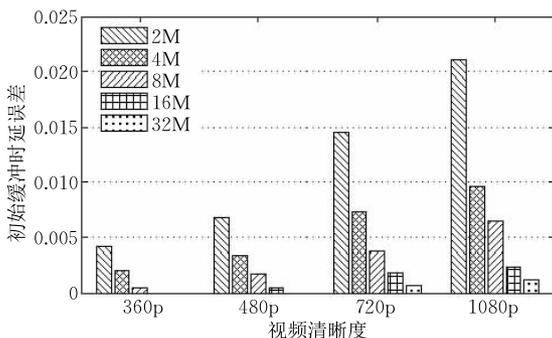


图 13 不同清晰度的初始缓冲时延评分误差

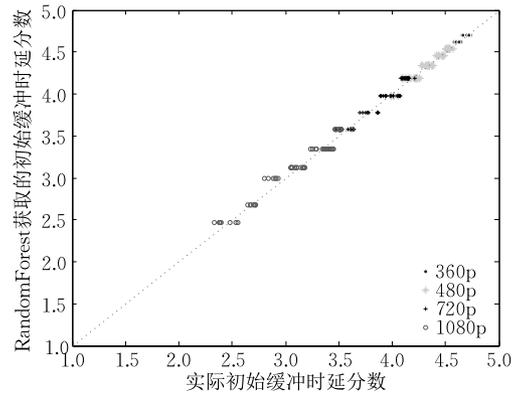


图 14 不同清晰度的视频“Cristiano Ronaldo—A Great Person”的初始缓冲时延评分误差

从图 13 可以看出,初始缓冲平均速度越大,初始缓冲因素平均误差越小.当初始缓冲速度为 2 Mbps,不同清晰度的误差介于 0.4%~2.1%,当初始缓冲速度为 32 Mbps 时,误差仅为 0.1%,因为初始缓冲时延越短,缓冲时延评估分值越高,误差相对较小.清晰度越高,平均误差越大,因为初始缓冲速度不变的情况下,清晰度越高缓冲时延越长,评估分值越低,误差相对较大.

### (3) 码率对卡顿相关的影响

卡顿占比具体反映当用户观看视频时总卡顿时间占总播放时间的比例.卡顿占比评估如式(3)所示,SR 表示卡顿时长占比,SR = 0.5%, 10%, 15%, 30% 分别对应卡顿相关值的 5, 4, 3, 2, 1. 然而,卡顿占比 = 卡顿时长和 / 观看时长,观看时长 = 文件大小 / 码率 + 卡顿时长和,因此卡顿因素也与码率相关.文献[1]研究发现,最大可承受卡顿时长为 3 s,因此,设置 0.5 s, 1 s, 1.5 s 和 3 s 这 4 种不同的卡顿时间进行比较,码率对卡顿相关的影响如图 15 所示.

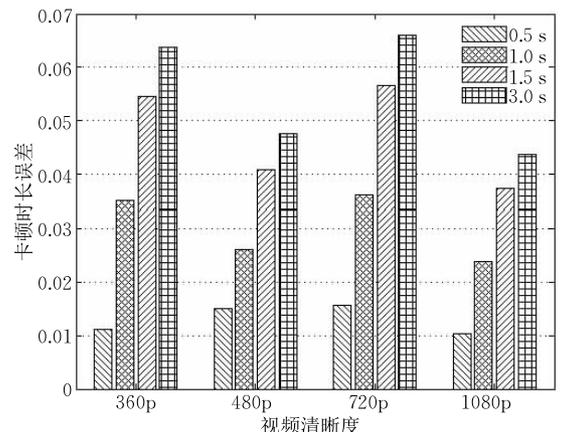


图 15 码率对卡顿因素的影响

从图 15 可以看出,当卡顿时长为 0.5 s 时,真实卡顿与估计卡顿的最小误差为 1%,当卡顿时长为

3 s 时,不同清晰度视频的误差介于 4.4%~6.7%。因为卡顿时长越长,基础分值越小,码率对卡顿影响的误差相对较大。

#### (4) 码率对 vMOS 的影响

码率对视频源质量、卡顿相关和初始缓冲感知的影响,最终影响 vMOS 评估,如式(4)所示。码率误差对 vMOS 的影响如图 16 所示。

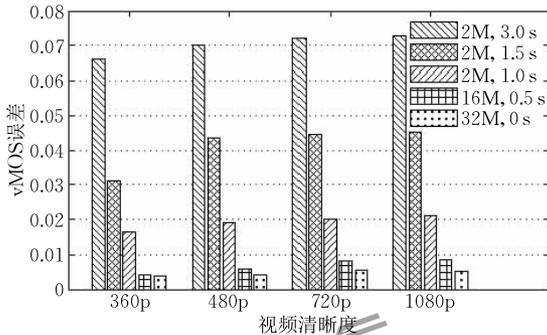


图 16 码率误差对 vMOS 的影响

从图 16 可以看出, vMOS 受码率的最终影响,初始缓冲平均速度越小,卡顿时长越长, vMOS 评估误差越大,最高平均误差为 7.2%,因为随着网络状况的优化, vMOS 分值变高,平均误差相对变小, vMOS 平均误差最小达到 0.4%。当前 vMOS 达到 4 为较好的状态( vMOS 为 1~5),一般介于 3~4,因此,最大误差引起的变化值介于 0.22~0.29,最小误差引起的变化值介于 0.01~0.02。

综合来看,基于机器学习的识别方法对 vMOSKPIs (视频源质量、卡顿相关和初始缓冲感知)影响相对较小,最终引起的 vMOS 误差也不大,可以很好的适用于加密 YouTube 视频 QoE 评估。

#### (5) 视频 vMOS 评估分布

基于 QoE 分析,采用 vMOS 评估多个移动网络环境下的 QoE 分布,如图 17 所示。基于这个评估,可以了解当前网络环境的视频服务质量。

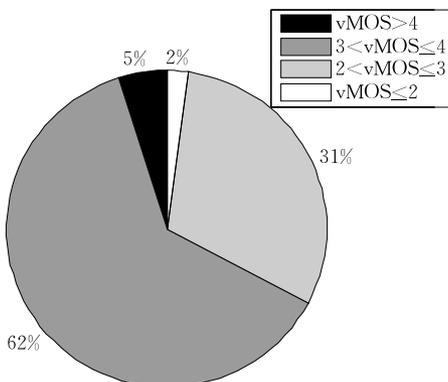


图 17 移动网络环境下 vMOS 分布

从图 17 可以看出,约 5% 的视频 vMOS 分值高于 4,说明这些视频的用户体验是较好的。62% 的视频 vMOS 分值介于 3~4,这些视频的用户体验是可接受的,31% 的视频 vMOS 分值介于 2~3,这些视频的用户体验较差的,而 2% 的视频 vMOS 分值低于 2,这些视频的用户体验非常差,当 vMOS 评分低于 3 时,将报告网络服务提供商和视频服务商,使其调整网络参数和传输模式改善视频服务体验。

## 6 结 论

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 HTTPS 加密传输,使得传统 DPI 方法无法获取视频大小、可播放时间等信息,最终导致原有的视频 QoE 评估方法失效。针对此问题,本文提出基于视频块特征的加密视频 QoE 参数识别方法:首先根据 SSL/TLS 协议握手过程中 ClientHello 数据包的未加密部分识别加密 YouTube 流量;然后根据视频流的前几个包的特征识别出不同传输模式,再根据视频块统计特征建立分类模式识别视频块的码率和清晰度;最后根据该视频参数和传输参数计算视频 QoE。鉴于机器学习码率标记是离散型数据,文中还分析了码率误差对视频 QoE 评估的影响。实验结果表明该方法可以有效用于加密视频 QoE 参数识别,具有较高的码率和清晰度识别性能,可以实现真实准确的视频 QoE 评估。下一步工作将结合用户主观感受改进 QoE 评估模型,并将 QoE 评估实时反馈网络服务提供商用于动态调整网络参数及网络故障诊断。

## 参 考 文 献

- [1] Hoßfeld T, Seufert M, Hirth M, et al. Quantification of YouTube QoE via crowdsourcing//Proceedings of the IEEE International Symposium on Multimedia (ISM) 2011, California, USA, 2011: 494-499
- [2] Mok R K P, Chan E W W, Luo X, et al. Inferring the QoE of HTTP video streaming from user-viewing activities//Proceedings of the 1st ACM SIGCOMM Workshop on Measurements Up the Stack, Toronto, Canada, 2011: 31-36
- [3] Nam H, Kim B H, Calin D, et al. Mobile video is inefficient: A traffic analysis. Nexus, 2013, (1): 1-5
- [4] Seufert M, Wamser F, Casas P, et al. YouTube QoE on mobile devices: Subjective analysis of classical vs. adaptive video streaming//Proceedings of the 2015 International Wireless Communications and Mobile Computing Conference (IWCMC). Dubrovnik, Croatia, 2015: 43-48

- [5] Seufert M, Egger S, Slanina M, et al. A survey on quality of experience of HTTP adaptive streaming. *Communications Surveys & Tutorials*, 2014, 17(1): 469-492
- [6] Nam H, Kim K H, Calin D, et al. YouSlow: A performance analysis tool for adaptive bitrate video streaming//*Proceedings of the 2014 ACM Conference on SIGCOMM*. Chicago, USA, 2014: 111-112
- [7] Seufert M, Wamser F, Casas P, et al. Demo: On the Monitoring of YouTube QoE in Cellular Networks from end-devices //*Proceedings of the 2015 Workshop on Wireless of the Students, by the Students, & for the Students*. Paris, France, 2015: 23-23
- [8] Wamser F, Seufert M, Casas P, et al. YoMoApp: A tool for analyzing QoE of YouTube HTTP adaptive streaming in mobile networks//*Proceedings of the 2015 European Conference on Networks and Communications (EuCNC)*. Paris, France, 2015: 239-243
- [9] Casas P, Seufert M, Schatz R. YOUQMON: A system for on-line monitoring of YouTube QoE in operational 3G networks. *ACM SIGMETRICS Performance Evaluation Review*, 2013, 41(2): 44-46
- [10] Nguyen T T T, Armitage G. A survey of techniques for internet traffic classification using machine learning. *Communications Surveys & Tutorials*, 2008, 10(4): 56-76
- [11] Bonfiglio D, Mellia M, Meo M, et al. Revealing skype traffic: when randomness plays with you. *ACM SIGCOMM Computer Communication Review*, 2007, 37(4): 37-48
- [12] Bernaille L, Teixeira R, Akodkenou I, et al. Traffic classification on the fly. *ACM SIGCOMM Computer Communication Review*, 2006, 36(2): 23-26
- [13] Bermolen P, Mellia M, Meo M, et al. Abacus: Accurate behavioral classification of P2P-TV traffic. *Computer Networks*, 2011, 55(6): 1394-1411
- [14] Paxson V. Empirically derived analytic models of wide-area TCP connections. *IEEE/ACM Transactions on Networking*, 1994, 2(4): 316-336
- [15] Zander S, Nguyen T, Armitage G. Self-learning IP traffic classification based on statistical flow characteristics//*Proceedings of the International Workshop on Passive and Active Network Measurement*. Boston, USA, 2005: 325-328
- [16] Paredes-Oliva I, Castell-Uroz I, Barlet-Ros P, et al. Practical anomaly detection based on classifying frequent traffic patterns //*Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. Orlando, Florida, USA, 2012: 49-54
- [17] Alshammari R, Zincir-Heywood A N. Unveiling Skype encrypted tunnels using GP//*Proceedings of the IEEE Congress on Evolutionary Computation*. Barcelona, Spain, 2010: 1-8
- [18] Bonfiglio D, Mellia M, Meo M, et al. Detailed analysis of skype traffic. *IEEE Transactions on Multimedia*, 2009, 11(1): 117-127
- [19] Chen K T, Huang C Y, Huang P, et al. Quantifying Skype user satisfaction. *ACM SIGCOMM Computer Communication Review*, 2006, 36(4): 399-410
- [20] Bar-Yanai R, Langberg M, Peleg D, et al. Realtime classification for encrypted traffic//*Proceedings of the International Symposium on Experimental Algorithms*. Naples, Italy, 2010: 373-385
- [21] Korczynski M, Duda A. Markov chain fingerprinting to classify encrypted traffic//*Proceedings of the 2014 INFOCOM*. Toronto, Canada, 2014: 781-789
- [22] Khakpour A R, Liu A X. An information-theoretical approach to high-speed flow nature identification. *IEEE/ACM Transactions on Networking*, 2013, 21(4): 1076-1089
- [23] Dubin R, Dvir A, Pele O, et al. Real Time Video Quality Representation Classification of Encrypted HTTP Adaptive Video Streaming-the Case of Safari. *arXiv preprint arXiv: 1602.00489*, 2016
- [24] Le H T, Nguyen H N, Pham Ngoc N, et al. A novel adaptation method for HTTP streaming of VBR videos over mobile networks. *Mobile Information Systems*, 2016, Article ID 2920850: 1-11
- [25] Lim Y, Kim H, Jeong J, et al. Internet traffic classification demystified: on the sources of the discriminative power//*Proceedings of the 6th International Conference*. Philadelphia, USA, 2010: 9
- [26] Kim H, Claffy K C, Fomenkov M, et al. Internet traffic classification demystified: Myths, caveats, and the best practices //*Proceedings of the 2008 ACM CoNEXT Conference*. Madrid, Spain, 2008: 11
- [27] Ramos-Munoz J J, Prados-Garzon J, Ameigeiras P, et al. Characteristics of mobile YouTube traffic. *Wireless Communications*, 2014, 21(1): 18-25
- [28] Yang Zhe, Li Ling-Zhi, Ji Qi-Jin, et al. Network traffic classification using decision tree based on minimum partition distance. *Chinese Journal on Communications*, 2012, 33(3): 90-102(in Chinese)  
(杨哲, 李领治, 纪其进等. 基于最短划分距离的网络流量决策树分类方法. *通信学报*, 2012, 33(3): 90-102)
- [29] Rissanen J. Stochastic complexity and the MDL principle. *Econometric Reviews*, 1987, 6(1): 85-102



**PAN Wu-Bin**, born in 1987, Ph. D. candidate. His current research interests include network security, network measurement and traffic classification.

**CHENG Guang**, born in 1973, Ph. D., professor, Ph. D. supervisor. His current research interests include network security, network measurement and behavior, future Internet security.

**WU Hua**, born in 1973, Ph. D., associate professor. Her current research interests include network security, network measurement and network management.

**XU Jian**, born in 1992, M. S. candidate. His current research interests include network measurement, traffic classification.

## Background

YouTube is one of the most popular and volume-dominant video streaming service in today's Internet. Almost 50% of the YouTube views are from mobile users, and this trend is expectedly increasing in near future. Understanding the user perception (i. e. QoE) on the performance and quality of the video streaming service is thus paramount for the video content provider YouTube and its content delivery network providers.

Since the client-side measurement needs the customers' cooperation to install some assessment application. And it cannot demonstrate that the decrease of QoE caused by the specific network links. In-network measurements (e. g. YOUQMON) have better coverage on end users. However, the resulting QoE has to be estimated from traffic characteristics typically via deep-packet inspection, which becomes infeasible after HTTP video streaming replaced by HTTPS. Recently, HTTPS adopted by major video content providers including YouTube and NetFlix to provision video services to mobile users with better protection on user privacy. Adaptive streaming is also commonly used as an effective means to enhance user QoE by dynamically adjusting bitrate suited to

current network condition.

The previous video quality assessment methods compute the bitrate from the video size and playback duration based on deep packet inspection or YouTube API. However, we are not able to get these parameters from the encrypted traffic when the QoE evaluation conducted in the network. Therefore, it requires with an urgency a DFI (deep flow inspection) technology which could be applied to encrypted video traffic.

This work was supported by the National High Technology Research and Development Program (863 Program) of China (2015AA015603), the Prospective Research Programs Future Internet of Jiangsu Province (BY2013095-5-03), the Six Talent Peaks of High Level Talents Project of Jiangsu Province (2011-DZ024), the Fundamental Research Funds for the Central Universities and the Research and Innovation Project for College Graduates of Jiangsu Province (KYLX15\_0118).

The research team has focused on network management and security for years, and more than ten papers in this domain have published in highly-ranked conferences and journals.