

# Impacto enlace ascendente YouTube 360° Live Video Streaming

Luis Roberto Jiménez<sup>(1)</sup>, Marta Solera Delgado<sup>(2)</sup>, Matías Toril Genovés<sup>(2)</sup>  
{lrjp, msolera, mtoril}@ic.uma.es

Dpto. de Ingeniería de Comunicaciones, ETSI Telecomunicación, Universidad de Málaga.

**Resumen**—At present, the most important service in mobile networks is video streaming, which is expected to generate 82 % of all IP traffic volume by 2021. YouTube is one of the most popular video services on the Internet. In March 2015, YouTube added 360° video streaming service, which provides users with panoramic view and allows them to freely control their viewing direction during video playback. It is getting popular due to the recent availability of commercial devices that support 360 video interactivity as smartphones/tablets and head-mounted displays. 360° live video streaming is very challenging due to high bandwidth requirements. To address this problem, the video quality is adjusted according to the user viewport prediction. High-quality video is only streamed for the user viewport while the rest of the frame is transmitted in lower quality, reducing the overall bandwidth consumption, but this solution is not sufficient for the high requirements demanded by users. Thus, the optimization of this service is the utmost importance for network operators. In this paper, a study of the uplink in the broadcast video live section and its impact on the quality of experience perceived by the end-user is presented. The results show video quality, end-to-end latency and MOS (Mean Opinion Score). The analysis has been carried out by collecting traces at network level for TCP/IP metrics, HTTP messages by HAR traces every 300s for video segment quality and end to end latency and user-level traces of each live 360° video session for S-KPIs (Service Key Performance Indicators).

## I. INTRODUCCIÓN

En la última década con los avances en la tecnología multimedia, su proliferación mediante internet [1] y el despliegue de futuras redes 5G, el interés sobre aplicaciones de emisión de video ha crecido exponencialmente ocasionando la aparición de nuevos casos de usos sobre redes móviles [2]. Paralelamente, los avances tecnológicos han aumentado las expectativas de los usuarios móviles, obligando a los operadores a cambiar la forma en que administran sus redes. En 2019, la emisión de video en directo ha superado el crecimiento de otros tipos de videos en línea [3] gracias al auge de aplicaciones para el entretenimiento como la retransmisión de eventos deportivos y culturales, la videovigilancia y la teleconferencia para reuniones y enseñanza a distancia. Además, el avance y el abaratamiento de las cámaras de video hace que los usuarios puedan realizar sus transmisiones en directo en formatos cada vez más complejos y sofisticados como el video en 360°. Básicamente, es un video panorámico grabado usando cámaras omnidireccionales [4]. Luego se proyecta en 2D utilizando una de las técnicas de mapeo disponibles (equiangular, cubo, piramidal, etc.) permitiendo a los usuarios mediante el ratón, una pantalla táctil de un teléfono o cascos de realidad virtual cambiar libremente la dirección de visualización del contenido, causando un efecto panorámico en la experiencia de video percibida [5]. Las principales plataformas de servicio, tales como *YouTube*, *Instagram* y

*Facebook*, emplean este tipo de emisiones, siendo el primero el líder indiscutible del mercado. Sin embargo, la distribución de videos 360 plantea nuevos desafíos. Independientemente del mapeo utilizado en una emisión de video en directo 360°, el cliente descarga la escena completa, aunque el usuario sólo visualice una parte que va entre los 90° a los 120° horizontales denominado puerto de visualización (*viewport*). Lo que se traduce en un uso ineficiente del ancho de banda de red al descargar datos que nunca se usan en la visualización real. Una de las soluciones empleadas para mejorar el rendimiento de red es la transmisión de video 360 adaptativo, en este caso, la dimensión espacial también se agrega al espacio de adaptación. En el cliente, se predice el puerto de visualización del usuario y se solicitan esos segmentos con mayor calidad que en el resto de la escena. En el servidor, se almacena el flujo de video original recibido y se transcodifica en segmentos de diferentes calidades de acuerdo al conjunto de *viewports* posibles. Por lo que en función de las condiciones de red y de la predicción de visualización se optimiza la descarga del video. Entender qué factores influyen y su impacto en los enlaces ascendente y descendente puede ser la clave para proporcionar una mayor calidad de experiencia (QoE) al usuario y proporcionar un esquema para diseñar algoritmos efectivos para optimizar el rendimiento de emisión de video en directo 360°.

Hasta la fecha, los diferentes trabajos publicados esbozan una serie de análisis y métodos para estimar la QoE a partir de diferentes indicadores de rendimiento del servicio (S-KPI), tomando en cuenta únicamente para el análisis, el enlace descendente desde el servidor de contenido al terminal de usuario. Prometedores estudios [6] [7] presentan modelos probabilísticos de predicción del *viewport* y optimización de la calidad de la imagen, basado sólo en la descarga de una porción visible de videos 360 según la predicción del movimiento de la cabeza y del estado del búfer del cliente.

En este artículo, se presenta un estudio del enlace ascendente en la sección emisora del video 360° en directo y su impacto en la calidad de experiencia percibida por el usuario final. El análisis detalla los S-KPIs utilizados para la consecución de la MOS, la calidad de cada segmento de video entregada y la latencia de extremo a extremo. El resto del artículo se estructura de la siguiente manera. La sección II describe la plataforma experimental. La sección III explica la metodología experimental. La sección IV muestra los resultados. Por último, la sección V, expone las conclusiones del trabajo.

## II. PLATAFOMA EXPERIMENTAL

La figura 1 muestra un diagrama de la plataforma utilizada para automatizar la recolección de mediciones. La plataforma está compuesta por dos módulos implementados sobre una red piloto LTE: un módulo de emisión (parte izquierda del diagrama) responsable de la emisión en directo de un video en 360° por la plataforma *YouTube Live*, y un módulo de medición (parte derecha del diagrama) responsable del descifrado, recolección y análisis de medidas. Ambos módulos utilizan como red de acceso una red piloto LTE. El módulo de emisión consta de un PC Dell OptiPlex (modelo 790 MT, procesador Quad Core i7 3.4 GHz, 8 GB de RAM, tarjeta de video integrada Intel HD Graphics 2000 Dynamic, disco duro de 500 GB) y una cámara de emisión en directo (modelo Samsung Gear 360°), conectada al PC vía USB 3.0. El software *Cyberlink Samsung Gear 360 Action* permite la creación y emisión del video 360° en tiempo real sobre la plataforma de *YouTube* desde un ordenador [8], mediante un evento previamente configurado en el gestor de emisión (*YouTube Studio*). Para la conexión a la red LTE se utiliza un dongle (modelo Huawei 4G LTE USB) con una *sim card* dada de alta en el gestor de red piloto LTE. Por otro lado, el módulo de medición consta de dos agentes de usuario, un emulador de red y una sonda de red. En el terminal móvil *Samsung Galaxy S9* se ejecuta el primer agente de usuario llamado *Network Performance Test (NPT)*, es una aplicación diseñada por Ericsson que además de imitar las interacciones de un usuario en una sesión de emisión de video en directo permite recopilar mediciones S-KPI (reporte NPT) [9]. En un PC (modelo Lenovo Thinkpad T560, procesador I5 8th Gen, 8 GB de RAM, tarjeta de video Intel HD Graphics 520 y disco duro 500 GB) se ejecuta el segundo agente de usuario que replica el comportamiento de un usuario en un terminal emulado *Samsung Galaxy S9* recopilando todos los mensajes descifrados y que está desarrollado mediante un script en Python, a partir de *Selenium Webdriver* y las herramientas de desarrollo de *google Chrome*. Para generar el reporte (reporte selenium), se extraen los parámetros referentes a la calidad del video recibida (*itag*) y la latencia de extremo a extremo de los mensajes HTTP que reportan medidas de QoE que se reciben con una frecuencia aproximada de 10s. Los agentes de usuario están conectados a la red piloto mediante una picocelda diferente a la utilizada por el módulo de emisión.

En un PC con Ubuntu 19.10 (*Eoan Ermine*) con procesador i5-750 a 3 GHz, 8 GB de RAM y dos tarjetas de red enlazadas mediante una tabla de enrutamiento, se ejecuta el emulador de red *NetEm* para modificar las condiciones de la red de forma controlada (por ejemplo, el caudal del enlace ascendente). Éste se coloca entre la picocelda del módulo de emisión y el núcleo de la red piloto LTE permitiendo estrangular el enlace ascendente del módulo de emisión (THRU\_UP). Las medidas de red a nivel de paquete se recopilan en la interfaz del emulador de red con acceso a Internet mediante una herramienta de captura estándar. Estos datos se procesan luego con una aplicación de monitoreo y análisis de tráfico (*network probe*) para extraer métricas básicas de QoS para cada conexión. Estos datos se utilizan para relacionar métricas de red con los S-KPI. Así en la plataforma se habilitan dos puntos principales de medida: la salida de emisión de video y

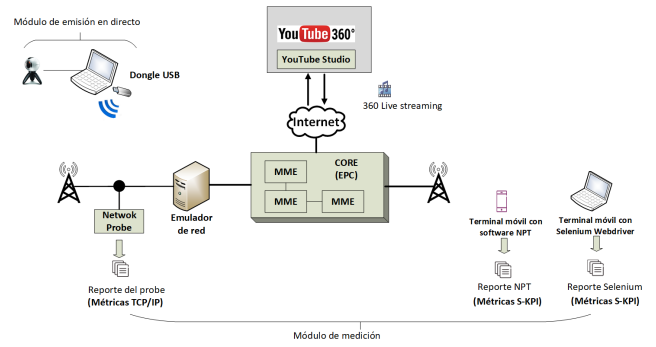


Fig. 1. Plataforma experimental de pruebas.

en recepción, el terminal de usuario. El primero se dedica a la recopilación de estadísticas de rendimiento de la transmisión de paquetes a nivel de red (métricas TCP/IP). Mientras que en el segundo se extraen los S-KPI necesarios para estimar la QoE de una sesión de video 360. Aquí, además, se descifran las cabeceras de los mensajes HTTP para la extracción de los parámetros *itag* y *latency*.

La red de acceso radio (red piloto LTE) está formada por 12 picoceldas de interior (modelo BTS3911B de Huawei) conectadas a un núcleo de red compacto (modelo eCNS600 de Huawei), de las que solamente dos se utilizan en estos experimentos, un servidor Wifi y un equipo que alberga a la herramienta de gestión de red (modelo RH5885 de Huawei). Todos estos elementos están conectados a través de un switch (modelo S5700 de Huawei).

## III. METODOLOGÍA EXPERIMENTAL

El objetivo del trabajo es identificar las variables del enlace ascendente en la sección de emisión de video, que impactan en la calidad de experiencia percibida por el usuario final, mediante el análisis de diversas trazas de sesión y red. Para ello, en primer lugar, se definen los S-KPIs más relevantes. En segundo lugar, se explica cómo se lleva a cabo la batería de pruebas.

### A. Definición de los S-KPI

La experiencia del usuario en emisiones en directo de *YouTube* se caracteriza por tres problemas básicos: el retardo inicial de reproducción del video (*Initial buffer time*), la detección de la reproducción del video (*rebuffering/stalling event*) y la duración de este estancamiento [10] [11]. Estos indicadores se obtienen a partir herramienta NPT: *Video start time*, *Stall time ratio* y *Freeze event rate* que además muestra el promedio temporal de la resolución de video en píxeles dada por su lado menor (altura o ancho dependiendo de si el video es horizontal o vertical respectivamente). Su cálculo consiste en una media ponderada, donde las diferentes resoluciones se promedian ponderadas por su tiempo acumulado durante la reproducción del video. Estos indicadores junto con la medida de QoE cuyo cálculo se basa en el modo 0 de la recomendación ITU-T P.1203.3 [12] que utiliza los eventos, la frecuencia y la duración recopilados por la herramienta permite calcular al NPT la MOS (*mean opinion score*) que predice el impacto de las variaciones de calidad de experiencia percibida por el usuario final en una emisión en directo de 360°.

TABLA I  
MAPEO DE *itag* PARA EMISIONES ADAPTATIVAS *YouTube Live 360°*. [13]

Valor <i>itag</i>	Resolución	Codificación	Bitrate (Mbits/s)
133	426 x 240	H.264	0.3 - 0.7
134	640 x 360	H.264	0.4 - 1
135	854 x 480	H.264	0.5 - 2
136	1280 x 720	H.264	1.5 - 4
137	1920 x 1080	H.264	3 - 6
264	2560 x 1440	H.264	9 - 18
266	3840 x 2160	H.264	13 - 34

Además, se proponen dos nuevos indicadores que miden la calidad de imagen de video y la latencia. El primero, la calidad de video ofrecida al usuario (*itag*), es entendida como el formato y resolución de cada segmento de video que compone el servicio *YouTube Live 360°*. Es un valor cuantitativo de calidad asignado a cada segmento de *Live 360° videostreaming* codificado, basado en su resolución, tasa de transmisión y técnicas de codificación. Se calcula como el máximo de los *itag* para un mismo segmento de video. La Tabla I muestra algunos valores de *itag* para diferentes resoluciones y formatos de emisiones en directo de video en *360°*.

El segundo parámetro, la latencia extremo a extremo, se define como la diferencia de tiempo entre el momento en que la cámara capta el evento y los espectadores lo visualizan. Cuanto menor sea la latencia menor será la cantidad de almacenamiento en búfer y por ende más rápido llegara el contenido al usuario final. Los servidores de *YouTube* destinados a emisiones en directo almacenan una pequeña cantidad de datos de la emisión para afrontar problemas debido a cambios de velocidad.

#### B. Batería de pruebas

Para identificar el impacto de los nuevos S-KPIs sobre el rendimiento de un servicio *Live 360° videostreaming* de *YouTube*, se realizaron baterías de pruebas durante el periodo comprendido del 1 al 30 de abril del 2020. Éstas consistieron en la emisión de un video *360°* mediante una cámara de emisión en directo. En el software gestor de la cámara se selecciona el perfil de video de 2560 x 1280 pixeles a 15 fps y usa la plataforma de *YouTube* como pasarela para la masificación del *Live video*. Una vez lanzado el *Live videostreaming*, se ejecuta de forma sincronizada el emulador de red y los clientes. Cada cliente tiene un tiempo de sesión del *Live 360° videostreaming* de 300s y se repite hasta finalizar la batería de pruebas. La capacidad del enlace ascendente se modifica mediante el emulador de red NetEm. Se estrangula el enlace a los siguientes valores: 0.5, 1, 2, 5, 10 y 20 Mbps. Cada configuración se mantiene durante 60 minutos. El tiempo total de una batería de pruebas es de 6 horas.

## IV. RESULTADOS

A continuación, se desglosan los resultados del análisis para los S-KPIs seleccionados específicos del servicio de emisión en directo *360°*, del que no se han publicado resultados similares anteriormente.

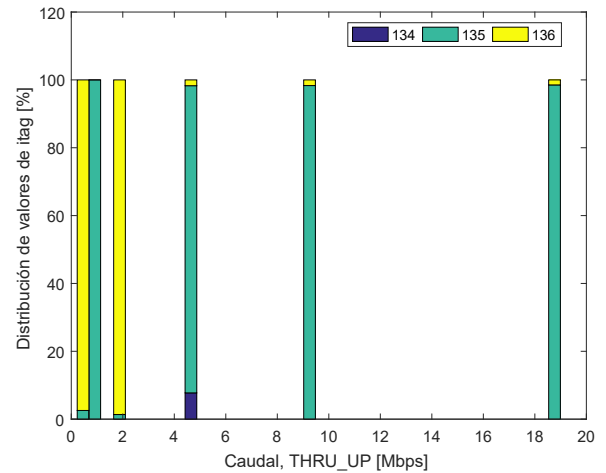


Fig. 2. Distribución de valores de *itag* frente tasa de datos disponible en el enlace ascendente.

#### A. Calidad de video

La figura 2 muestra un diagrama de barras apiladas con la distribución de *itag* en función del caudal del enlace ascendente. En abscisas se representa el valor medio de los valores de *Throughput* (THRU\_UP) extraídos en el punto de emisión del video. En ordenadas se muestra el porcentaje de *itags* de cada clase dentro de cada columna, extraído mediante el procesamiento de los archivos HAR capturados en el agente de usuario emulado. Primero, se puede observar que sólo aparecen 3 valores de *itags* (134, 135, 136), con 134 y 136 las resoluciones mínima y máxima. Para valores de estrangulamiento de 5, 10 y 20 Mbps se observa que el valor predominante de *itag* es 135 que se corresponde con una resolución de 854 x 480 y una tasa de video de 0.5 a 2Mbps. En el caso de 5Mbps, también aparecen sesiones que utilizan resoluciones más bajas (*itag* 134). Como resultado, el estrangulamiento del enlace ascendente no afecta directamente a la calidad de video recibida por el cliente. Cuando se limita la capacidad del enlace ascendente por debajo de los 2Mbps, los resultados no son fiables. Un estudio más exhaustivo, no mostrado aquí, muestra que prácticamente el 100% de las sesiones para 0.5 y 1Mbps no permiten descargar en directo el contenido. Para 2 Mbps, este porcentaje se reduce al 30%. Por lo que los valores de *itag* que se muestran en la figura 2 corresponden a los primeros cuadros y no resultan representativos.

#### B. Latencia

La figura 3 muestra la función de distribución de la latencia extremo a extremo para valores de estrangulamiento del canal ascendente de 5, 10 y 20 Mbps. Para un valor de 5 Mbps, la curva muestra los valores de latencia más elevados con un promedio de 45.09s. Es en el primer tercio de la conexión donde se obtienen los valores más pequeños que oscilan entre 30s y 40s para aumentar paulatinamente a medida que progresa la reproducción. Las curvas para 10 y 20 Mbps muestran valores de latencia más contenidos con valores medios de 20.21s y 24.92s respectivamente. No se muestran en la gráfica las curvas para 0.5, 1 y 2 Mbps porque la latencia presenta valores anómalos. En estos casos, el comportamiento

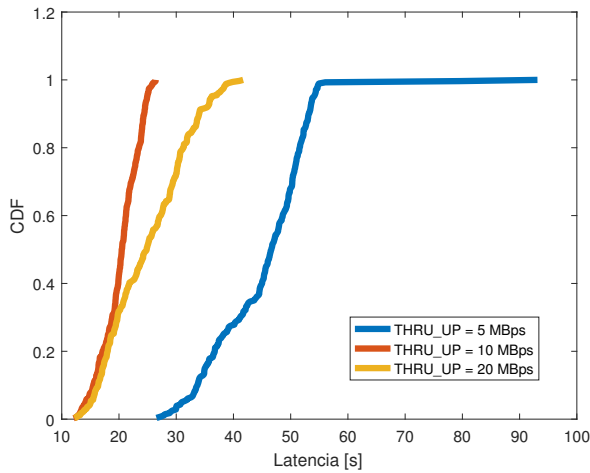


Fig. 3. Valores de latencia frente tasa de datos disponible en el enlace ascendente.

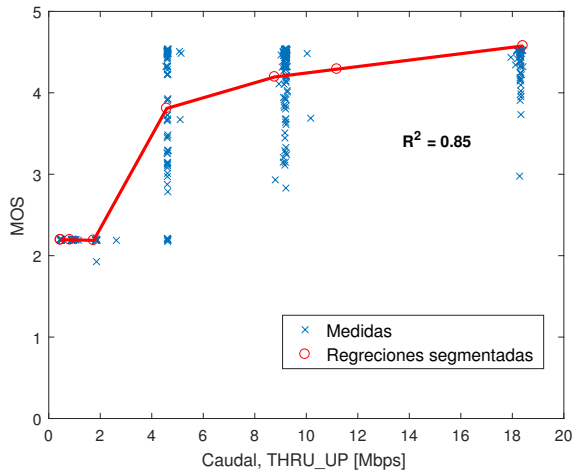


Fig. 4. QoE estimada frente tasa de datos disponible en el enlace ascendente.

es creciente con el tiempo, comenzando con valores de unos 200s al comienzo de la conexión y alcanzando los 3600s después de una hora de reproducción.

### C. QoE

La figura 4 muestra un grafo de dispersión de la MOS estimada por el terminal NPT para una sesión del *Live 360° videostreaming* en función del caudal ascendente (THRU\_UP). En la gráfica cada punto corresponde a una sesión de 300s para un mismo estado de la red. La curva de regresión segmentada que mejor ajusta los datos se superpone, lo que resulta en un coeficiente de determinación de  $R^2 = 0,85$ .

En la figura, se observa que los puntos se agrupan en columnas que coinciden con las limitaciones de ancho de banda (valores de limitación en NetEm). La curva de regresión muestra cómo el QoE disminuye a medida que THRU\_UP disminuye. Específicamente, un THRU\_UP de 4 Mbps da como resultado un MOS cercano a 3.6, por debajo del máximo de 4.5.

## V. CONCLUSIONES

En este artículo se ha presentado un estudio del enlace ascendente en la sección emisora del video 360° en directo y su impacto en la QoE percibida por el usuario final. El análisis detalla la calidad de cada segmento de video entregada, la latencia de extremo a extremo y los S-KPIs utilizados para la consecución de la MOS. El análisis se ha realizado mediante la recolección de trazas a nivel de red para las métricas TCP/IP, captura de trazas HAR cada 300s para la calidad del segmento de video y la latencia, y trazas a nivel de usuario para el cálculo de la MOS.

El análisis de los datos recogidos muestra que en la emisión en directo de video 360° los segmentos de videos descargados por el cliente tienen un *itag* predominante de 135 (resolución de 854 x 480) y 0.5-2Mbps. En el video 360° el cliente obtiene todo el video, incluidas las partes visibles e invisibles, permitiéndose seleccionar la dirección de visualización mientras reproduce el video. Los resultados demuestran que el enlace ascendente no influye en la calidad de los segmentos de video descargados por el cliente. En cambio, si afecta a la latencia y QoE, ambos, son inversamente proporcionales al estrangulamiento. De los resultados, se deduce la importancia del estudio de la latencia extremo a extremo y su impacto sobre la QoE. Esto justifica la necesidad de nuevas vías de desarrollo para estimarla desde métricas TCP/IP.

## AGRADECIMIENTOS

Este trabajo ha sido financiado por el Ministerio de Ciencia, Innovación y Universidades (Proyecto TEC2015-69982-R y RTI2018-099148-B-I00).

## REFERENCIAS

- [1] Sandvine, "Global internet phenomena report," Tech. Rep., 2018.
- [2] NGMN Alliance, "5G White paper," 2015.
- [3] P. Jonsson, S. Carson *et al.*, "Ericsson mobility report," *Ericsson AB, Technol. Emerg. Business, Stockholm, Sweden, Tech. Rep. EAB-17*, vol. 5964, 2019.
- [4] Samsung. Introducing the 360 round vr camera. [Online]. Available: <https://www.samsung.com/us/business/products/mobile/virtual-reality/360-round-vr-camera/>
- [5] S. Hollister. (2015) YouTube's ready to blow your mind with 360-degree videos. [Online]. Available: <https://gizmodo.com/youtubes-ready-to-blow-your-mind-with-360-degree-videos-1690989402>
- [6] L. Xie, Z. Xu, Y. Ban, X. Zhang, and Z. Guo, "360probdash: Improving qoe of 360 video streaming using tile-based http adaptive streaming," in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, pp. 315–323.
- [7] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan, "Optimizing 360 video delivery over cellular networks," in *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, 2016, pp. 1–6.
- [8] CyberLink. Introducing gear 360 actiondirector. [Online]. Available: <https://www.cyberlink.com/learning/gear-360-actiondirector/606/introducing-gear-360-actiondirector>
- [9] Ericsson, *S-KPI Benchmarking 20.3 NPT User Guide*, 2020.
- [10] T. Hofffeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz, "Quantification of youtube qoe via crowdsourcing," in *Multimedia (ISM), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 494–499.
- [11] P. Casas, A. Sackl, S. Egger, and R. Schatz, "YouTube & Facebook Quality of Experience in mobile broadband networks," in *2012 IEEE Globecom Workshops (GC Workshops)*, 2012, pp. 1269–1274.
- [12] R. ITUTP, "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport-quality integration module," *International Telecommunication Union*, 2017.
- [13] YouTube, "Live encoder settings, bitrates, and resolutions," <https://support.google.com/youtube/answer/2853702?hl=en>, accessed on Nov. 2018.