



UNIVERSIDAD
POLITECNICA
DE VALENCIA



Máster Universitario
en Tecnologías, Sistemas y
Redes de Comunicaciones

Nuevas técnicas para optimizar el tráfico de red utilizando Big Data

Autor: Álvaro Daniel Liébana Carrascosa

Director: Jaime Lloret Mauri

Resumen

Con el auge del Big Data, el enfoque hacia el análisis del tráfico de red NTMA supone un gran reto. Cuando se trabaja con redes, es bastante común el uso de software especializado para analizar el tráfico que se genera en internet, el cual nos permite, ente otros, el diagnostico de problemas de red, análisis de protocolos y extracción de información de las tramas donde viajan los datos. Sin embargo, hay información que no se puede ver directamente desde los campos de la trama como por ejemplo la calidad de un video de internet que está reproduciendo un usuario. En este trabajo, obtendremos una forma para identificar y clasificar esta información.

Abstract

With the rise of big data, the approach to network traffic analysis NTMA poses a great challenge. When you work with networks, it's fairly common the use of software specialized in analysing the traffic that the internet generates, which allows us, among other things, the diagnosis of network problems, analysis of protocols and extraction of information from the data frames.

But nevertheless, there's information that you can't see directly from the fields of the data frame like for example the quality of a video that a user is playing. In this research paper, we will obtain a way of identifying and clasifying this information.

Autor: Álvaro Daniel Liébana Carrascosa, email: alliecar@epsa.upv.es

Director: Jaime Lloret Mauri, email: jlloret@com.upv.es

Fecha de entrega: 30-11-2020

ÍNDICE

I.	INTRODUCCIÓN	7
I.1	Introducción	7
I.1.1	Plataformas OTT	7
I.1.2.	¿Qué es TCP y UDP?	8
I.1.3.	Resolución del video	10
I.1.4.	¿Qué es el bitrate?	10
I.1.5.	Programas utilizados	11
I.1.6.	Proceso de captura.....	11
I.2.	Objetivos	12
I.3.	Precedentes del proyecto.....	12
I.4.	Estructura del proyecto.....	13
II.	Estado del arte	14
III.	propuesta del proyecto	16
IV.	Desarrollo del metodo de detección de calidad de video	21
V.	Exposicion de los resultados obtenidos.....	22
VI.1.	Características teóricas.....	22
VI.1.1.	Twitch	22
VI.1.2.	YouTube	22
VI.1.3.	Netflix	23
VI.1.4.	TV online	23
VI.1.5.	Facebook.....	23
VI.2.	Características reales	24
VI.2.1.	Twitch.....	24
VI.2.2.	YouTube	27
VI.2.3.	Netflix	29
VI.2.4.	TV online	29
VI.2.5.	Facebook.....	30

VI.	Conclusiones	32
VI.1	Cumplimiento del objetivo.....	32
VI.2	Conclusiones sobre el proyecto	33
VI.3	PROBLEMAS ENCONTRADOS Y SOLUCIONES.....	33
VI.4	FUTURAS LINEAS DE TRABAJO	34
VII.	REFERENCIAS	35

ÍNDICE DE FIGURAS

Fig. 1. Clasificación de tráfico en 2019 por la empresa Sandvine	7
Fig. 2. Comparativa protocolos de transmisión TCP, UDP y QUIC.....	9
Fig. 3. Paso 1 para iniciar captura en TShark.....	11
Fig. 4. Paso 2 para iniciar captura en TShark.....	12
Fig. 5. Paso 3 para iniciar captura en TShark.....	12
Fig. 6. Categorías de la NTMA.....	16
Fig. 7. Conocimiento descubierto a partir de los datos en bruto.....	17
Fig. 8. Captura de trafico de una transmisión	21
Fig. 9. Captura retransmisiones en la plataforma Twitch con calidad 1080p60	25
Fig. 10. Captura retransmisiones en la plataforma Twitch con calidad 720p60	25
Fig. 11. Captura retransmisiones en la plataforma Twitch con calidad 720p	26
Fig. 12. Captura retransmisiones en la plataforma Twitch con calidad 480p	26
Fig. 13. Captura retransmisiones en la plataforma YouTube con calidad 1080p	27
Fig. 14. Captura retransmisiones en la plataforma YouTube con calidad 720p	28
Fig. 15. Captura retransmisiones en la plataforma YouTube con calidad 480p	28
Fig. 16. Captura retransmisiones en la plataforma Netflix con calidad alta	29
Fig. 17. Captura retransmisiones en plataformas de Tv online con calidad 720p.....	30
Fig. 18. Captura retransmisiones en la plataforma Facebook con calidad 1080p.....	30
Fig. 19. Captura retransmisiones en la plataforma Facebook con calidad 720p.....	31
Fig. 20. Captura retransmisiones en la plataforma Facebook con calidad 480p.....	31

INDICE DE TABLAS

Tabla 1. parámetros teóricos de bitrate por calidad en Twitch.....	22
Tabla 2. parámetros teóricos de bitrate por calidad en YouTube.....	23
Tabla 3. parámetros teóricos de bitrate por calidad en Facebook.	23
Tabla 4. parámetros reales de bitrate por calidad en Twitch.....	27
Tabla 5. parámetros reales de bitrate por calidad en YouTube.	29
Tabla 6. parámetros reales de bitrate por calidad en Facebook.....	31
Tabla 7. Tabla final bitrate por calidad.	32

I. INTRODUCCIÓN

I.1 INTRODUCCIÓN

Cuando se trabaja con redes, es bastante común el uso de software especializado para analizar el tráfico que se genera en internet, el cual nos permite, ente otros, el diagnostico de problemas de red, análisis de protocolos y extracción de información de las tramas donde viajan los datos. Estas herramientas nos proporcionan gran información sobre el tráfico, y nos permite aplicar filtros cuando trabajamos con grandes cantidades de datos. Sin embargo, hay gran parte de la información que no podemos extraer observando solamente los campos de las tramas como, por ejemplo, la calidad a la que se está reproduciendo un video.

En la actualidad, el tráfico de contenido multimedia es vital en la vida de muchas personas consumiendo gran parte de su tiempo de ocio, y no es de extrañar en la era de internet en la que vivimos, con altas velocidades al alcance de mucha gente. Gracias al trabajo de Sandvine, empresa de telecomunicaciones especializada en inteligencia de red, podemos cifrar el consumo multimedia que hay en el tráfico de internet a nivel mundial ya que realizan un reporte anual el cual podemos consultar libremente. En dicho reporte correspondiente al año 2019 [1], el consumo video ha supuesto el 60.6% del tráfico en internet (Fig.1), teniendo un incremento del 2.9% respecto al reporte del año 2018 [2].

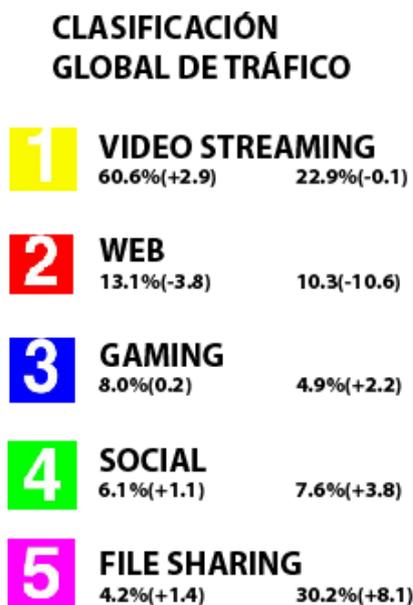


Fig. 1. Clasificación de tráfico en 2019 por la empresa Sandvine

I.1.1 PLATAFORMAS OTT

La mayor parte de este contenido en video viene de las aplicaciones y plataformas Over-The-Top, llamadas plataformas OTT. Estas son aplicaciones que ofrecen contenido de video a través de internet sustituyendo la forma tradicional de verlo a través de televisión, ya sea por cable o por satélite.

Gracias a ellas podemos reproducir el contenido no solo en ordenadores, sino también en dispositivos móviles como Smartphones o tablets o incluso en la televisión mediante hardware externo (Chromecast, Amazon Fire stick TV, etc) o ya sea por la propia naturaleza Smart de la TV.

En este trabajo, hemos evaluado algunas de las plataformas más populares que ofrecen contenido vía streaming. Con el concepto streaming, nos referimos a cualquier contenido de medios, ya sea en vivo o grabado (bajo demanda) que podemos disfrutar a través de internet y en tiempo real. El funcionamiento del streaming es sencillo, a grandes rasgos, los datos de video que se van a transmitir se organizan en paquetes secuenciales de datos, sin embargo, a diferencia de las descargas tradicionales, estos archivos que descargas se eliminarán progresivamente tras ser reproducidos por el dispositivo que estés utilizando. Entre las plataformas analizadas se encuentran, Netflix, YouTube, Twitch, Facebook y Amazon Prime Video. Como veníamos diciendo anteriormente, el objetivo es obtener con un alto grado de acierto la calidad a la que se está reproduciendo el video.

1.1.2. ¿QUÉ ES TCP Y UDP?

Cuando hablamos de contenido multimedia, ya sea en streaming o bajo demanda, este está asociado a un protocolo elegido para realizar la comunicación en la capa de transporte del modelo TCP/IP. A continuación, resumiremos sus características y nombraremos sus principales usos.

TCP: El protocolo TCP (Protocolo de Control de Transmisión) tiene como parte fundamental, que los datos lleguen correctamente al destinatario a través de las tramas (más convenientemente llamados segmentos), sin errores, y en orden. En caso de perderse alguno de estos segmentos, el protocolo TCP inicia su retransmisión para garantizar que los datos lleguen sin errores y sin necesidad de la intervención de la capa de aplicación. Este protocolo es especialmente utilizado en aplicaciones de video bajo demanda debido a su fiabilidad, pero no es tan recomendado para retransmisiones en vivo o videollamadas puesto que en caso de que se detecte en el destino que los paquetes no están en orden, se pedirá el paquete o paquetes restantes mientras el receptor se encuentra en un estado de espera, algo que en videos bajo demanda no es un gran problema, pero sí que lo es en retransmisiones interactivas o videollamadas, entre otros. Sin embargo, gran parte de las plataformas que ofrecen retransmisiones en directo, utilizan TCP debido a que retrasan unos segundos dicha transmisión para cargar el buffer y tener más fluidez a la vez que se solventa en gran medida el problema. El protocolo TCP ofrece soporte a gran cantidad de protocolos de la capa de aplicación como, por ejemplo, HTTP, HTTPS, POP3, SMTP, etc.

UDP: El protocolo UDP (Protocolo de Datagramas de Usuario) nos permite el envío de datagramas sin la necesidad de establecer previamente una conexión siendo únicamente necesario un socket abierto por parte del destino para que los datagramas que vienen del origen sean aceptados. UDP es un protocolo no orientado a conexión, es decir, al contrario que ocurre en el protocolo TCP

el cual contiene una fase para establecer una conexión, en este caso, se envían sin previo aviso. Además, no proporciona ningún tipo de control de flujo, por lo que tiene que hacer uso de la capa de transporte ya que no tiene ningún mecanismo de reenvío de datagramas perdidos. Por último, carece de control de congestión para aliviar dicha congestión en la red, así como tampoco control de errores. Debido a todas estas carencias, las garantías para la transmisión de la información se implementan en capas superiores y este protocolo es usado principalmente en DHCP y DNS donde es más importante la rapidez que la fiabilidad. En cuanto a su uso en aplicaciones de video, es altamente utilizado para transmisiones de audio y video en vivo donde lo más importante es la rapidez de la transmisión, y no tanto su calidad.

QUIC: Durante los análisis de tráfico de la plataforma YouTube se ha detectado que este circulaba a través de UDP. Teóricamente debería utilizar TCP dados los beneficios mencionados anteriormente y al ser, en su mayoría, contenido bajo demanda y, de hecho, es lo que estaba sucediendo hasta hace no mucho. Sin embargo, de la mano de Google para mejorar las limitaciones de UDP, nació el nuevo protocolo HTTP/3 más conocido como QUIC [3] (Quick UDP Internet Connections) por sus siglas en inglés, dónde la IETF también está dentro del proyecto [4].

Esta nueva versión hará que se deje de utilizar el protocolo TCP en HTTP y ya se está empezando a incorporar, el cual tenía como previsión de llegada oficial este 2020.

Como podemos apreciar en la tabla comparativa, QUIC se deshace de las constantes conexiones entre emisor y receptor, eliminando también el intercambio de datos con el que se establece el cifrado TLS, por tanto, QUIC únicamente envía los datos. Este, además, reducirá en gran medida la latencia de la conexión. Existen otros beneficios que se pueden consultar en su blog oficial de desarrollo del protocolo [5].

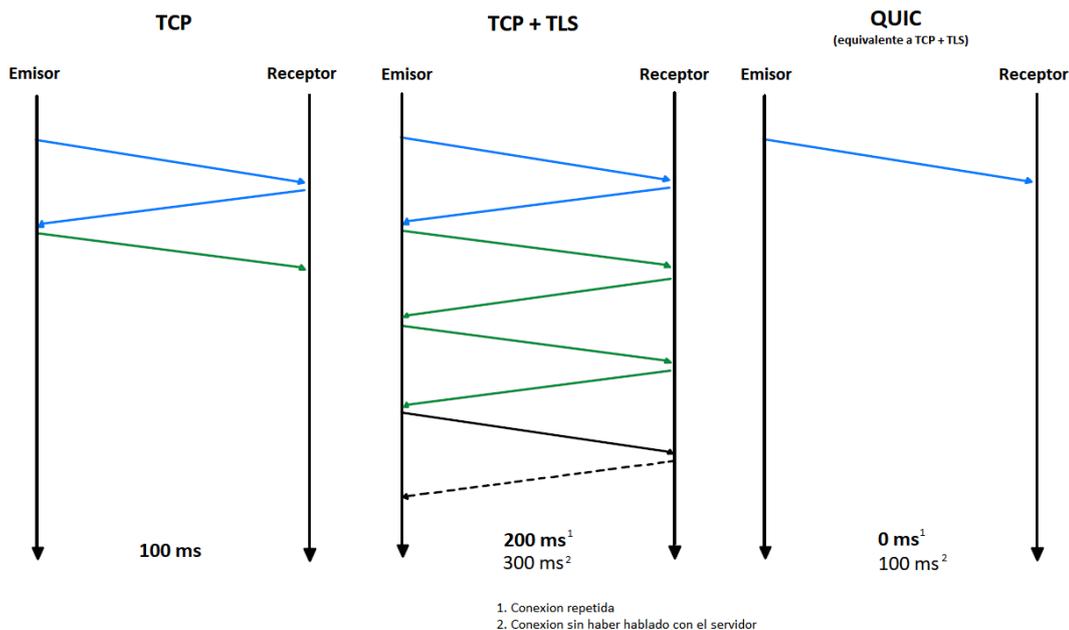


Fig. 2. Comparativa protocolos de transmisión TCP, UDP y QUIC.

1.1.3. RESOLUCIÓN DEL VIDEO

Al mismo tiempo que la tecnología, las velocidades cada vez son más elevadas y, por tanto, nos permiten utilizar resoluciones mucho mayores. Actualmente la gran mayoría de dispositivos admiten resoluciones Full HD, sin embargo, poco a poco se va instaurando la que será la nueva resolución de referencia: estamos hablando del 4K.

Hace años que se escucha hablar sobre esta resolución, pero ahora ha llegado para quedarse pues no solo dispositivos de alta gama la poseen sino un amplio espectro de precios en el mercado dispone de ella.

La tendencia de la resolución avanza cada vez con mayor velocidad pues aún sin haber realizado la gran mayoría el salto generacional de resolución, ya hay gente que habla incluso de 8K y hay televisores que montan paneles con esta resolución.

La importancia de dicha tendencia en este trabajo reside en que aun podemos encontrar gran parte de contenido multimedia actual en resoluciones HD (sobre todo en países menos avanzados), pero que, junto con los nuevos avances tecnológicos en componentes de ordenador se ira eliminando hasta establecerse una resolución mínima más aceptable como es la resolución Full HD.

Sin embargo, puesto que aún se utilizan resoluciones menores a Full HD, también se han estudiado en este trabajo.

1.1.4. ¿QUÉ ES EL BITRATE?

De igual forma que la resolución o el formato del vídeo, otro factor importante que influye en la calidad del video es el bitrate. El bitrate es la tasa de datos, es decir, la cantidad de datos almacenados por cada segundo que se está reproduciendo. Cuanto más alta sea esta velocidad, mayor será la calidad del material a visualizar. Sin embargo, aumentar este valor también dará lugar a un video más pesado y en caso de que el destinatario no posea una conexión a internet de alta calidad es posible que sufra parones durante la reproducción del video.

Otro factor clave para tener en cuenta es la compresión del video. Para comprimir se utilizan algoritmos especializados que consiguen ofrecer una alta calidad con tasas de bits más modestas, aunque en numerosas ocasiones, los resultados no son los esperados, y se realizan compresiones agresivas que empeoran enormemente la calidad del video que es subido a plataformas como YouTube, Vimeo, etc.

1.1.5. PROGRAMAS UTILIZADOS

Para este proyecto se han utilizado los siguientes programas:

Wireshark: Esta herramienta es un potente sniffer de red que intercepta todo el tráfico que circula a través de ella y lo convierte en un formato que pueda ser leído por las personas. Gracias a esta herramienta se pueden utilizar filtros para identificar los paquetes que más le puedan interesar al cliente que lo utiliza.

TShark: Esta variante está integrada en wireshark y es una forma de realizar capturas de red sin la necesidad de un entorno gráfico, únicamente utilizando la ventana de comandos. En ella se realizarán las capturas y se exportarán al formato CSV que posteriormente abriremos con Microsoft Excel.

Microsoft Excel: Software con interfaz gráfica dividida en celdas que permite realizar gran cantidad de cálculos usando los datos de las tablas ya que posee diferentes funciones personalizadas. En él se realizarán los procesos necesarios para conseguir los graficas comparativas del trabajo.

1.1.6. PROCESO DE CAPTURA

El método para realizar la captura del tráfico de video es mediante el programa Wireshark, aunque hemos utilizado su herramienta sin interfaz gráfica TShark para agilizar el proceso mediante la línea de comandos. Además, TShark es capaz de detectar, leer y escribir los mismos archivos de captura que los que son compatibles con Wireshark.

```
C:\Program Files\Wireshark>tshark -Dir
1. \Device\NPF_{EF578F63-3397-42E5-BF30-9F4E05FDE0DE} (Conexión de área local* 10)
2. \Device\NPF_{6C3D7BC3-F11B-4998-98E1-6B71E91AA93C} (Conexión de área local* 8)
3. \Device\NPF_{3C452CEE-47F8-40FA-862E-09CD12F6906E} (Ethernet)
4. \Device\NPF_{82EB0D39-6C0A-4CCB-A821-DC8F0AB279D3} (Wi-Fi)
5. \Device\NPF_{0C2FE8E9-467C-469A-BFEF-1B5E0914ED08} (Conexión de área local* 1)
6. \Device\NPF_{E324509A-02DC-4FFA-B726-C3DA31755C38} (Conexión de red Bluetooth 2)
7. \Device\NPF_{9A201C5D-BCA5-458D-8F7E-03E71911A045} (Conexión de área local* 2)
8. \Device\NPF_{DC4BA836-D678-4473-BC34-27CC7EDE7029} (Conexión de área local* 9)
9. \Device\NPF_{51B6ADE9-287F-4CCA-BA68-F36538BC7437} (Npcap Loopback Adapter)
10. \Device\NPF_Loopback (Adapter for loopback traffic capture)
11. \Device\NPF_{F92BE789-05D2-4A9F-9195-8E3D563B7A2E} (Ethernet 2)
12. \\.\USBPcap1 (USBPcap1)

C:\Program Files\Wireshark>
```

Fig. 3. Paso 1 para iniciar captura en TShark.

Una vez iniciadas las transmisiones se ha esperado alrededor de 1 minuto para que se estabilice y mientras tanto consultaremos la lista de interfaces (Fig.3) para seleccionar en el siguiente paso cual desearemos utilizar para la captura como se aprecia en la. Posteriormente se ha procedido a iniciar la captura limitando los paquetes a 100.000 seleccionando en este caso la interfaz es ethernet (interfaz nº3 de la lista de la Fig.3) para una aproximación mayor (Fig.4).

```
C:\Program Files\Wireshark>tshark -i 3 -w origen.pcap -c 100000
```

Fig. 4. Paso 2 para iniciar captura en TShark.

Una vez tenemos el archivo de captura con formato pcap vamos a convertirlo al formato csv (mediante el propio TShark) con expresión de la Fig.5.

```
C:\Program Files\Wireshark>tshark -r origen.pcap -T fields -e frame.number -e frame.len -e frame.time_relative -e eth.s  
rc -e eth.dst -e ip.src -e ip.dst -e ip.proto -e tcp.srcport -e tcp.dstport -e udp.srcport -e udp.dstport -E header=y -  
E separator=, -E quote=d -E occurrence=f >destino.csv
```

Fig. 5. Paso 3 para iniciar captura en TShark.

Con ella lo que estamos haciendo es extraer los parámetros que nos interesan y establecer el método de separación por comas para que se abra correctamente en Excel y de esta forma nos permitirá manejar los datos con la finalidad de realizar operaciones y poder visualizar los resultados de una forma más comparativa creando graficas.

1.2. OBJETIVOS

El principal objetivo de este trabajo es conseguir detectar potencialmente la resolución del video que está visualizando un usuario desde cualquier parte de internet, principalmente de las páginas de contenido multimedia más populares, analizando su tráfico. Esta información no se puede ver directamente desde los campos de la trama por ello, con este método, podremos tener una idea más detallada del tráfico.

1.3. PRECEDENTES DEL PROYECTO

Se han encontrado algunos trabajos relacionas concretamente con el propósito de este proyecto. En [6] encontramos la creación de nuevas herramientas para monitorear la calidad de la transferencia de video con una arquitectura cliente-servidor donde recopila estadísticas de la calidad del video y las envía al servidor. Además, con el paso del tiempo también se están empezando a realizar más investigaciones para aprender sobre plataformas como YouTube [7] y [8]. Este trabajo sigue esa línea intentando obtener más información del tráfico de este tipo de servicios que están en auge.

I.4. ESTRUCTURA DEL PROYECTO

El desarrollo del trabajo inicia en el apartado II, donde se hablará sobre la situación actual del análisis de tráfico, así como las diferentes investigaciones para obtener información adicional.

Posteriormente, en el apartado III, se realiza la propuesta del proyecto con una explicación detallada de las fases necesarias para conseguir los objetivos deseados.

En la sección IV, se explica detalladamente el método para detectar los flujos de contenido streaming y como diferenciarlo de otro tipo de flujo que pueda parecer, pero que en realidad no lo sea.

En el apartado V se exponen los resultados obtenidos representados con graficas para comprar las diferencias entre las plataformas seleccionadas.

Finalmente, en el apartado 6, se podrán encontrar tanto las conclusiones, como los problemas encontrados durante el proyecto y futuras líneas de trabajo aplicables a este proyecto.

II. ESTADO DEL ARTE

Con el auge de Big Data, se pueden hacer frente a situaciones en las que el volumen, la velocidad, la veracidad y la variedad de datos suponen desafíos para permitir la extracción del valor de estos datos. Uno de los principales usos del Big Data está enfocado al análisis del tráfico de red (NTMA por sus siglas en inglés) ya que la supervisión y el análisis de este sector es una tarea complicada por los enormes volúmenes que se manejan y los cuales cada vez van aumentando con mayor velocidad debido al incremento en las velocidades de los sistemas de transmisión.

Para administrar y monitorear la red es necesario clasificar el tráfico para identificar los servicios que generan este tráfico [9]. Tiempo atrás los servicios de internet solían identificarse simplemente inspeccionando los números de puerto TCP/UDP. Como hemos comentado en varias ocasiones durante este proyecto, la clasificación del tráfico ya no es sencilla por razones como el aumento de la cantidad de servicios ofrecidos, sin embargo, el factor más complicado es la necesidad de identificarlos observando la poca información que permanece visible en los paquetes, ya que gran parte de los servicios se ejecutan sobre protocolos de encriptación (un claro ejemplo es HTTPS sobre TCP).

En [10] se presenta un estudio de medición de transmisión de video con tasa de bits adaptativa para un para un evento en vivo, impulsada por la popularidad de la transmisión de video en vivo como las plataformas vistas en este proyecto.

El compromiso con la satisfacción del usuario surge de la relación entre la calidad de experiencia (QoE) y la calidad del servicio (QoS). Mientras los proveedores de servicio abogan por una mejor calidad de experiencia, los proveedores de red prefieren enfocarlo hacia la calidad del servicio. Para medir los parámetros de QoS y entender la relación entre ambas, se ha realizado el siguiente estudio [11].

Gran parte de las investigaciones relacionadas en este campo intentan conseguir y aprender más parámetros que nos hagan entender mejor el comportamiento de la red y de la interacción del ser humano en ella. Por ejemplo, en [12] se analiza el impacto de las métricas de calidad de servicio como el almacenamiento en búfer, las métricas de calidad de la red (por ejemplo, la velocidad de datos de la capa física), el contenido del video (como su duración), etc.

Otros estudios, como [13] se centran en medir y analizar la QoE estudiando la satisfacción del usuario en diversas aplicaciones (como YouTube, Firefox, etc) relacionándolas con mediciones de QoS del host final (RTT, jitter, etc). En [14] se realizó un estudio para medir el QoE en las transmisiones de video. Se calcularon la puntuación de opinión medio (MOS) en videos que comienzan reproduciéndose en buena calidad, pero con el paso del tiempo se degradan y también para videos que comienzan en mala calidad, pero con el paso del tiempo mejoran. Determinaron que

los usuarios tenían una reacción más positiva cuando la calidad del video empezaba siendo deficiente y mejoraba con el tiempo.

También se han realizado estudios que pretenden obtener las tendencias de los contenidos que circulan por el portal de YouTube y saber el seguimiento por parte de la sociedad con el fin de utilizar estos datos para establecer modelos de negocio [15]. En [16] podemos ver otro enfoque de aprendizaje automático basado en transmisiones el cual trata de analizar el estancamiento de las sesiones de transmisión en YouTube en tiempo real a partir del tráfico cifrado de la red. También encontramos estudios relacionados con el análisis de conjuntos de datos que vienen de YouTube por medio de MapReduce ofrecido por la herramienta Hadoop [17].

El próximo paso de la NTMA es el análisis de datos, en [18] podemos ver uno de los principales métodos de análisis a grandes conjuntos de datos de la NTMA, aunque hay otros escenarios diferentes como en [19] y [20]. Sin embargo, uno de los grandes problemas relacionados con NTMA es la falta de datos públicos y relevantes que limitan de cierta manera la reproducibilidad donde solo algunas contribuciones como [21] donde dejan disponibles como código abierto tanto su código, como su conjunto de datos y su demostración. Sin embargo, esta práctica no suele ser habitual, lo que limita en cierto modo el punto de partida de muchas investigaciones que podrían ser más escalables.

En este proyecto, aportaremos nuestra visión para la detección de la calidad de los videos que reproducen los usuarios en internet.

III. PROPUESTA DEL PROYECTO

El gran volumen de datos que se mueve en el tráfico de redes aumenta a una velocidad cada vez mayor, y es necesario que estos datos sean capturados y analizados a altas tasas y, en muchas ocasiones, en tiempo real. El análisis de estos datos, que provienen de diferentes fuentes, nos abre horizontes a nuevas formas de utilizar la información que podemos extraer de la red. Esta información es extraída de los grandes volúmenes de datos gracias a las técnicas de minería de datos, y constantemente surgen nuevas formas para utilizar la información extraída en favor tanto de los usuarios como de las grandes empresas.

Anteriormente hemos mencionado la NTMA y su estrecha relación con el Big data. Sin embargo, dentro de la NTMA se distinguen diferentes categorías como podemos observar en la Fig.6.

Categoría	Descripción
Predicción de tráfico	Mantenimiento y planificación de redes.
Clasificación de tráfico	Categoriza e identifica flujos de tráfico de diferentes fuentes.
Gestión de fallos	Predice y aísla fallos y comportamientos indeseados en la red.
Seguridad de la red	Protege la red, reacciona a actividades maliciosas y ataques en general.

Fig. 6. Categorías de la NTMA

A pesar de que todas ellas nos ofrecen un entorno interesante de investigación, en este proyecto vamos a centrarnos en la clasificación de tráfico.

La clasificación de tráfico es crucial para administrar y monitorear la red identificando información sobre los servicios. Esta información nos puede ofrecer un conocimiento de los requisitos de la red y su impacto en el rendimiento. Hay numerosos enfoques propuestos para realizar la clasificación de tráfico, y en este proyecto optado por desarrollar una detección la información de calidad de los videos de diferentes plataformas de contenido multimedia.

Estas plataformas quieren ofrecer una gran experiencia calidad para el usuario, sin embargo, aún existen muchos casos de destinatarios de estos servicios que tienen una conexión relativamente mala, que no se ajusta a sus necesidades. Sin embargo, los proveedores de servicios no pueden detectar estas deficiencias en los servicios de los usuarios que experimentan dichos problemas.

Cuando visualizamos un video de cualquier lugar de internet, estamos descargando los datos en la calidad que nuestra conexión nos permite, y si tenemos una mala conexión se irá reduciendo hasta que podamos abarcar una reproducción del video sin necesidad de parones. Cuando el video que está siendo reproducido reduce su resolución, también lo hace la cantidad de datos que tenemos que descargar, y esa cantidad es fácilmente visible si utilizamos los filtros necesarios.

Aunque hay muchas páginas donde se ofrecen contenido de videos, se han elegido las que constituyen gran parte de este tráfico, ya sea bajo demanda o en vivo como son: Twitch, Youtube, Netflix, Facebook, y páginas de canales de televisión donde se puede visualizar esta desde el ordenador.

El proceso a seguir desde que se obtienen los datos de las capturas de tráfico es el que se puede observar en la Fig.7. Una vez tenemos los datos en bruto, identificaremos y seleccionaremos las variables más relevantes de estos datos, como pueden ser el tiempo del paquete, el puerto y la longitud de la trama, etc. Tras esta fase de selección entramos en la etapa de preprocesamiento en la cual se aplicarán operaciones básicas para preparar los datos que no están en el rango esperado. Por ejemplo, este es el caso del tiempo del paquete que tenemos que pasarlo a segundos pues el abrir el archivo no se da correctamente. Tras esta etapa ya tenemos los datos preparados para el siguiente bloque. En esta primera etapa del segundo bloque tenemos la tarea de buscar características útiles a los datos, es decir en ella se realizará el primer cribado o filtro donde se diferenciará el tráfico del video del que no lo es y posteriormente se calculará según el puerto el bitrate del flujo.

Por último, se visualizarán e interpretarán los patrones descubiertos consolidando el conocimiento y abriendo la posibilidad a incorporar lo aprendido en otros sistemas.

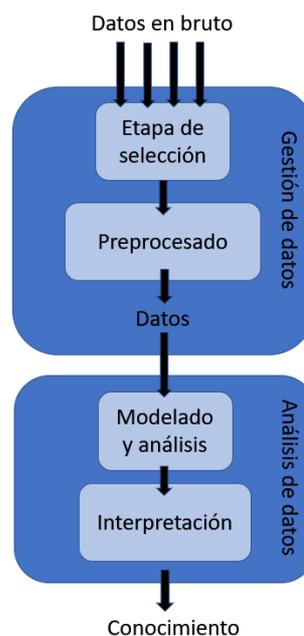


Fig. 7. Conocimiento descubierto a partir de los datos en bruto

En nuestro caso concretamente, lo primero que vamos a realizar es una captura de tráfico mientras se visualizan una gran cantidad de videos variando sus calidades. El objetivo es establecer un patrón en todas estas capturas de tráfico. El análisis se ha realizado por medio de la interfaz Ethernet con el fin de tener una lectura más precisa y sin pérdida de paquetes.

Una vez tenemos los archivos de las capturas, los convertiremos de PCAP a formato CSV para que puedan ser abiertos por el programa Excel, que es el que vamos a utilizar para proporcionar las gráficas a partir de este tráfico. Este paso se realiza por medio de TShark como se explica en el apartado I.6.

Teniendo ya los datos de las capturas listos para su manipulación, vamos a aplicar un filtro que organice todo el tráfico que queremos analizar por conexiones IP emisor y receptor que sucedan a través de un mismo puerto. Una vez tenemos todo el tráfico organizado habrá una serie de flujos que sean los que concentren mayor cantidad de información. Esta información en nuestro caso la podemos ver a simple vista desde Wireshark, consultaremos las capturas para ver cuál es el puerto con mayor flujo y usarlo para realizar las gráficas.

En un entorno que no sea Wireshark, tras organizar los flujos por direcciones IP y puerto, se sumará el campo de longitud de la trama, se seleccionarán los que sean potencialmente de video por la gran cantidad de bytes que se han intercambiado entre emisor y receptor. Esta es una primera criba que se realiza para posteriormente calcular el bitrate de los que cumplen dicho requisito ya que normalmente lo que ofrece mayor consumo de datos es la visualización de videos y las descargas.

Con estos datos se procede a calcular el bitrate, en nuestro caso este cálculo se ha realizado en Excel. Para ello hemos creado una función fraccionada en columnas y automatizada por la funcionalidad de Macros de Excel que permite realizar todos los cálculos automáticamente en el orden que se hayan programado. Nuestra función, recorre todo el bloque de datos y selecciona los que corresponden al puerto que hemos detectado que es el que concentra el mayor flujo de datos de la captura. Una vez los tenemos, sumamos todos los campos de longitud de tramas de los paquetes que son de dicho puerto y calculamos el bitrate con la duración la captura, que, ya que solo hemos estado consumiendo ancho de banda sin realizar otra tarea, el tiempo de visualización corresponda el tiempo que ha durado la captura. Sin embargo, en otro entorno el tiempo utilizado sería el que corresponda a la resta del ultimo paquete de la comunicación con el primero. De esta forma obtenemos el bitrate. El Código de la macro, se observa a continuación:

```
1. Sub Calculo_completo()  
2. '  
3. ' Calculo_completo_Macro  
4. '  
5.  
6. '  
7.  
8. Range("M3:M100001,M2").Select
```

```

9.      Range("M2").Activate
10.     Selection.FormulaR1C1 = "=RC[-10]/1000000000"
11.     Range("N2").Select
12.     ActiveCell.FormulaR1C1 = "1"
13.     Range("N3").Select
14.
15.     ActiveWindow.ScrollRow = 1
16.     ActiveCell.FormulaR1C1 = "10"
17.     Range("N4").Select
18.
19.     Range("P2:P100001").Select
20.     Selection.FormulaR1C1 = "=IF(RC[1]="""",NA(),RC[1])"
21.
22.     ActiveWindow.SmallScroll Down:=-10
23.     Range("Q3:Q100001").Select
24.     Selection.Formula2R1C1 = _
25.         "=IFERROR(IF((FLOOR(INDIRECT(""M"&(ROW()-
1)    1)),R2C14))=(FLOOR(INDIRECT(""M"&ROW()),R2C14)),""",IF(ROUNDDOWN(RC[-
4]    4),0)=0,"""",IF(MOD(FLOOR(RC[-4],R2C14),R2C14)=0,SUM(OFFSET(RC[-15],,-(ROW()-
(MATCH(ROUNDDOWN(INDIRECT(""M"&ROW()),0)-
R2C14,R2C18:R100001C18,0)))),"""")),SUM(OFFSET(RC[-15],,-((ROW()-
(MATCH((INDIRECT(""R"&(ROW()))-R2C14),R2C18:R100001C18,1))))))"
26.     Range("R1").Select
27.     ActiveCell.FormulaR1C1 = "0"
28.     Range("R2").Select
29.     ActiveCell.FormulaR1C1 = "0"
30.     Range("R3").Select
31.
32.     Range("R3:R100001").Select
33.     Selection.Formula2R1C1 = _
34.         "=IF((ROUNDDOWN(INDIRECT(""M"&(ROW()-
1)    1),0))=(ROUNDDOWN(INDIRECT(""M"&ROW()),0)),""0"",ROUNDDOWN(INDIRECT(""M"&ROW()
),0))"
35.
36.     Range("S2:S100001").Select
37.     Selection.Formula2R1C1 = "=IF(RC[-9]=INDIRECT(""0"&2),RC[-17],""")"
38.
39.     Range("T100001").Select
40.     Range(Selection, Selection.End(xlUp)).Select
41.     ActiveWindow.SmallScroll Down:=-40
42.     Range("T3:T100001").Select
43.     Selection.Formula2R1C1 = _
44.         "=IFERROR(IF((FLOOR(INDIRECT(""M"&(ROW()-
1)    1)),R2C14))=(FLOOR(INDIRECT(""M"&ROW()),R2C14)),""",IF(ROUNDDOWN(RC[-
7]    7),0)=0,"""",IF(MOD(FLOOR(RC[-7],R2C14),R2C14)=0,SUM(OFFSET(RC[-1],,-(ROW()-
(MATCH(ROUNDDOWN(INDIRECT(""M"&ROW()),0)-
R2C14,R2C18:R100001C18,0)))),"""")),SUM(OFFSET(RC[-1],,-((ROW()-
(MATCH((INDIRECT(""R"&(ROW()))-R2C14),R2C18:R100001C18,1))))))"
45.
46.     Range("V2").Select
47.     ActiveWindow.SmallScroll Down:=-30
48.     ActiveCell.FormulaR1C1 = "=(SUM(R2C23:R100001C23)*8)/R100001C13"
49.     Range("V3").Select
50.
51.     Range("V3:V100001").Select
52.     Selection.FormulaR1C1 = _
53.         "=IF(R[-1]C[2]="""",NA(),(SUM(R2C23:R100001C23)*8)/R100001C13)"
54.
55.     Range("W2:W100001").Select
56.     Selection.FormulaR1C1 = "=IF(RC[-13]=R2C15,RC[-21],""")"
57.
58.     Range("X3:X100001").Select
59.     Selection.Formula2R1C1 = _
60.         "=IFERROR(IF((FLOOR(INDIRECT(""M"&(ROW()-
1)    1)),R3C14))=(FLOOR(INDIRECT(""M"&ROW()),R3C14)),""",IF(ROUNDDOWN(RC[-
11]   11),0)=0,"""",IF(MOD(FLOOR(RC[-11],R3C14),R3C14)=0,SUM(OFFSET(RC[-5],,-(ROW()-

```

```
(MATCH(ROUNDDOWN(INDIRECT("M"&ROW()),0)-  
R3C14,R2C18:R100001C18,0)))),""))),SUM(OFFSET(RC[-5],,,-((ROW())-  
(MATCH((INDIRECT("R"&ROW()))-R3C14),R2C18:R100001C18,1)))))"  
61.  
62. End Sub
```

Una vez se ha calculado el bitrate de todos los contenidos visualizados, se procede a graficarlos uno a uno para ver que intervalos se dan en cada una de las calidades que se han seleccionado y estudiar si se da algún patrón entre ellas. Las columnas que se graficarán será la M que corresponde al tiempo en segundos y la V que corresponde a la media de bits de la comunicación. Si es el caso, como veremos más adelante en este proyecto, se establecerá como criterio para detectar potencialmente la resolución de un video que se está reproduciendo desde internet.

Además, se compararán estos intervalos con los datos que ofrecen algunas de estas plataformas sobre el bitrate que ofrecen, identificando así la veracidad de su información. Normalmente estas páginas sufren compresiones muy fuertes que alteran la buena calidad de los videos reduciendo significativamente la calidad de la grabación original. En [22] se puede observar el deterioro que sufre un video tras varias resubidas encadenadas, hasta hacerse completamente imposible de entender su contenido.

Como hemos comentado anteriormente, este tipo de técnicas puede desembocar en un mayor conocimiento de la interacción de los usuarios en la red por parte de las grandes empresas proveedoras de servicios y, gracias a estos conocimientos, utilicen dicha información para ofrecer mejores servicios al usuario final.

IV. DESARROLLO DEL METODO DE DETECCIÓN DE CALIDAD DE VIDEO

En este apartado vamos a explicar cómo identificamos y diferenciamos los datos de tráfico de video, del resto.

Para identificar una alta probabilidad de que el flujo es de video hemos puesto varios filtros para diferenciarlos de otro tipo de tráfico como puede ser la navegación por páginas web, descargas u otro tipo de tareas que podamos realizar por internet.

Como hemos mencionado anteriormente, el bitrate es un factor que varía en función de la calidad de reproducción del video. Sin embargo, para obtenerlo tenemos que realizar una serie de cálculos básicos. Para ello lo primero es detectar todo el tráfico que va de una IP origen a otra IP destino por el mismo puerto. Con esta información, sumaremos el campo de longitud de cada segmento o datagrama (en caso de utilizar el protocolo UDP) para obtener el total de bytes de la comunicación. Una vez tengamos este dato y, utilizando el tiempo de duración de la comunicación, que podemos conocer haciendo la resta del tiempo exacto en el que se envió el primer segmento y el ultimo, calcularemos la tasa de bits por segundo o bitrate.

Ethernet · 22		IPv4 · 48		IPv6 · 4		TCP · 34		UDP · 38							
Address A	Port A	Address B	Port B	Packets	Bytes	Packets A → B	Bytes A → B	Packets B → A	Bytes B → A	Rel Start	Duration	Bits/s A → B	Bits/s B → A		
192.168.0.11	61325	212.166.133.173	443	37.640	46M	4.009	372k	33.631	46M	2.767830	111.5999	26k	3339k		
192.168.0.11	52949	216.58.211.246	443	157	162k	31	3805	126	159k	2.696690	1.1499	26k	1106k		
192.168.0.11	59710	172.217.17.14	443	106	91k	22	6346	84	85k	2.111458	27.7381	1830	24k		
192.168.0.11	63947	216.58.211.246	443	90	81k	32	8582	58	73k	49.189362	0.2079	330k	2818k		
192.168.0.11	52948	216.58.209.65	443	65	57k	24	1980	41	55k	2.701472	0.7223	21k	610k		

Fig. 8. Captura de tráfico de una transmisión

Para diferenciarlo del resto de tráfico es sencillo pues la navegación ente paginas no consume tal cantidad de bytes. Sin embargo, existe la posibilidad de que haya una cifra similar por diversos factores, algo que podría llevar a confundir la predicción. Para evitar estos casos hemos fijado unos parámetros mínimos para empezar a comprobar el bitrate. Basados en numerosas pruebas realizadas, se calculará el bitrate de aquellas conexiones que superen los 20 MB en un periodo de 120 segundos. El bitrate también tendrá un máximo para evitar flujos de descarga de archivos grandes. Este filtro determinara si el flujo es potencialmente un video y dependiendo del bitrate, cuál es su calidad.

Por ejemplo, en la Fig.8 hemos marcado una comunicación que no es el flujo principal del video, pero que tiene un bitrate que podría ser confuso. Sin embargo, no supera ni los segundos establecidos ni el tamaño de 20 MB, por tanto, se determina que no es un flujo de video.

V. EXPOSICION DE LOS RESULTADOS OBTENIDOS

En este apartado se mostrarán los resultados de las lecturas de algunas de las plataformas más populares de videos tanto de retransmisiones en directo como bajo demanda, para fijar cuáles serán los umbrales para determinar la calidad del video.

VI.1. CARACTERÍSTICAS TEÓRICAS

Algunas páginas, brindan información oficial sobre el bitrate que recomiendan utilizar para sus transmisiones o videos. Esta información les sirve para optimizar sus algoritmos de compresión y, por tanto, al utilizar un bitrate mayor del recomendado, ellos lo ajustarán a dicho rango.

Sin embargo, otras, como el caso de Netflix, no hay datos oficiales con los que poder comparar los obtenidos de su contenido.

VI.1.1. TWITCH

Twitch es la plataforma de retransmisión en vivo más grande del mundo. Desde sus comienzos han estado especializados en un contenido enfocado a los videojuegos, sin embargo, en ella se ofrecen todo tipo de contenidos diferentes como música, vlogs, podcast, etc. Su repercusión es de tal magnitud, que, según el informe trimestral de streamlabs [23], ha sido con diferencia la plataforma preferida para los creadores de contenido durante el estado de alarma.

Según la página oficial [24] los bitrates recomendados para transmitir en las distintas calidades en Twitch son los de la Tabla 1.

Resolución	Bitrate
1080p60	6000 kbps
1080p30	4500 kbps
720p60	4500 kbps
720p30	3000 kbps

Tabla 1. parámetros teóricos de bitrate por calidad en Twitch.

Las emisiones en directo posteriormente son resubidas a la propia página web, con el bitrate original de la transmisión.

VI.1.2. YOUTUBE

Esta plataforma bien conocida por todos también tiene su variante para retransmisiones de videojuegos en vivo aprovechando el auge constante de este sector.

De igual manera los datos oficiales recomendados [25] para la página de YouTube se resumen en la Tabla 2.

Resolución	Bitrate
1080p60	4500-9000 kbps
1080p30	3000-6000 kbps
720p60	2250-6000 kbps
720p30	1500-4000 kbps
480p	500-2000 kbps

Tabla 2. parámetros teóricos de bitrate por calidad en YouTube.

En este caso, tanto las emisiones como los videos subidos a la plataforma deberían tener el mismo rango de bitrate para cada calidad de video especificado.

VI.1.3. NETFLIX

Hemos incluido en los análisis, a una de las plataformas de películas y series más conocida pues maneja una gran cantidad de tráfico hacia su contenido.

Sin embargo, como se comentaba al inicio del apartado II, al ser una plataforma privada donde los usuarios no pueden subir contenido, los datos de bitrate y compresión no son públicos. Sin embargo, analizaremos su bitrate real más adelante en el apartado III.2.

VI.1.4. TV ONLINE

De igual forma que sucede con el servicio anterior, dado que no hay datos oficiales que utilizan en sus retransmisiones, únicamente se analizarán sus valores reales del apartado III.2.

VI.1.5. FACEBOOK

Siguiendo la misma estrategia que la plataforma YouTube, la red social Facebook abrió un apartado dedicado a las retransmisiones, enfocada en gran medida al sector de los videojuegos.

Según la página oficial [26] los bitrates recomendados para transmitir en las distintas calidades se pueden consultar en la Tabla 3.

Resolución	Bitrate
1080p60	6000 kbps
1080p30	4500 kbps
720p60	4500 kbps
720p30	3000 kbps

Tabla 3. parámetros teóricos de bitrate por calidad en Facebook.

VI.2. CARACTERÍSTICAS REALES

Tener los datos oficiales de los parámetros que se usan en estas plataformas es importante. Sin embargo, a la hora de usar esta información para investigación no basta con tener esos parámetros teóricos, pues pueden estar alejados de la realidad. Por ello, siguiendo con los pasos de los apartados I.6 y II obtendremos la información del video que llega al consumidor del contenido de cada plataforma.

Tras las tablas del apartado anterior, se realizará una comparativa para esclarecer cuanto se acercan a los valores reales obtenidos durante la visualización del contenido.

VI.2.1. TWITCH

Cuando analizamos numerosas emisiones en directo a mayor calidad (Fig. 9), podemos apreciar que hay 2 valores de bitrate dominantes: 6600 kbps es el más común junto con 8800 kbps, en ellos se concentran la mayoría de los videos analizados. Como bitrate menos usado estaría el valor de 5800 kbps aproximadamente el cual únicamente se ha encontrado en una ocasión de todas las retransmisiones analizadas. Según los análisis no suele haber diferencias entre contenido, ya sea videojuegos, blogs, podcasts, etc. Sin embargo, en su página oficial recomiendan ciertos tipos de resolución en función del contenido de juego que se esté emitiendo, aunque como decimos en la práctica no se llevan a cabo estas recomendaciones por diversos factores individuales del creador de contenido.

A continuación, se muestran diferentes graficas con los resultados obtenidos:

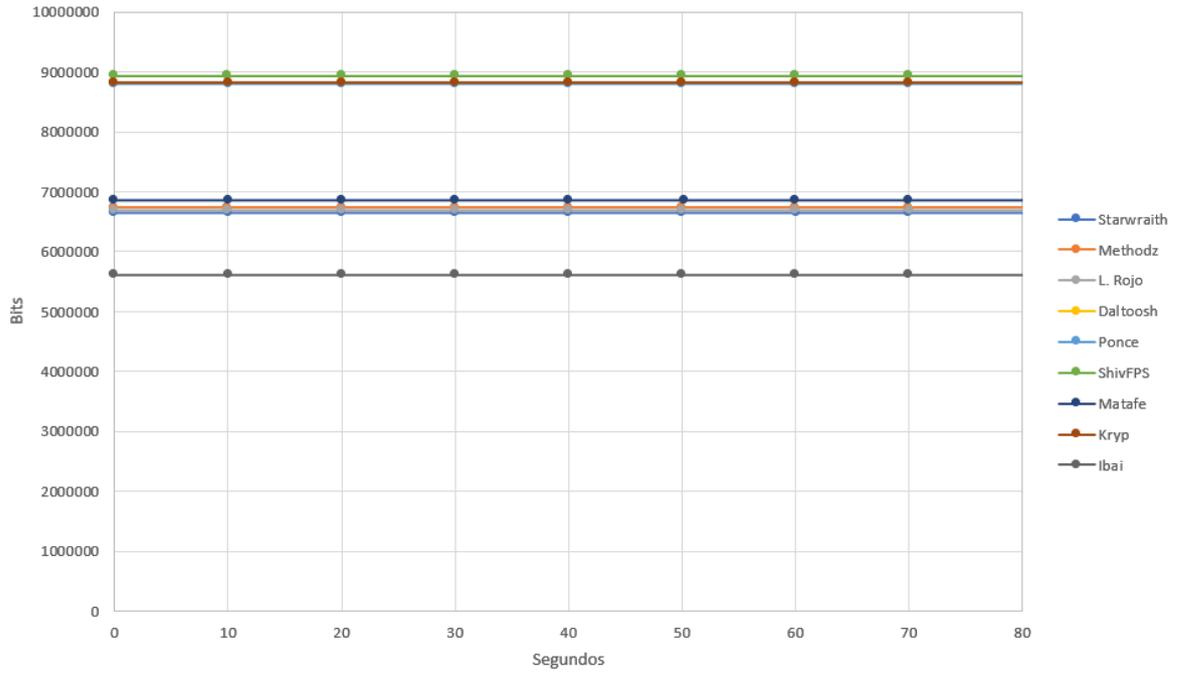


Fig. 9. Captura retransmisiones en la plataforma Twitch con calidad 1080p60

Para la calidad de 720p a 60 frames por segundo (Fig.10) y 30 frames (Fig.11) los valores se concentran en dos rangos bien diferenciados entre sí, de 3000-4000 kbps, y de 2000-3000 kbps.

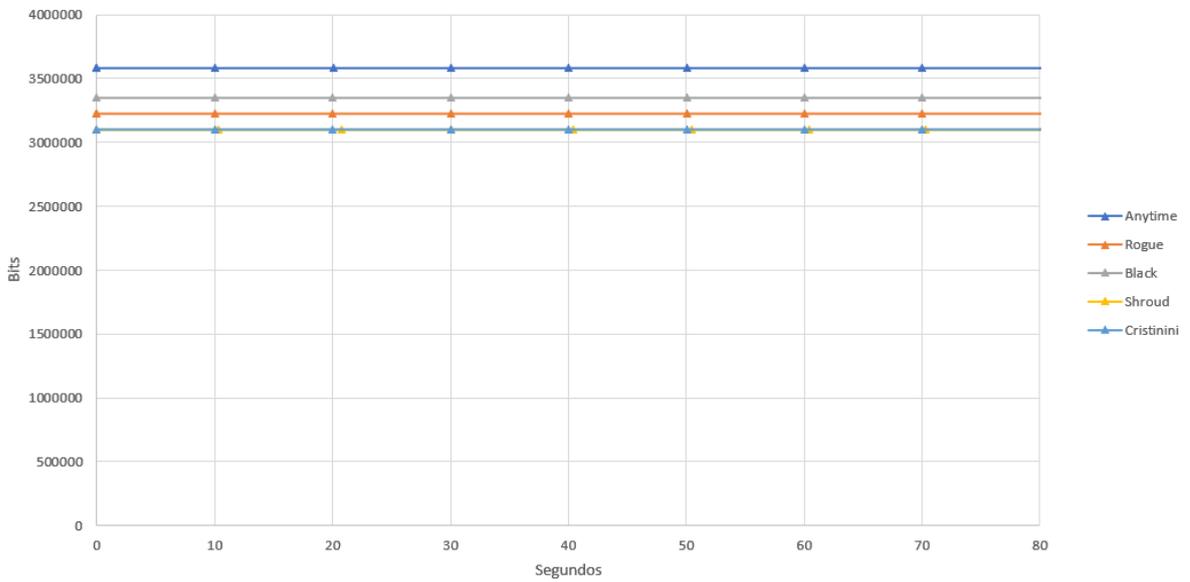


Fig. 10. Captura retransmisiones en la plataforma Twitch con calidad 720p60

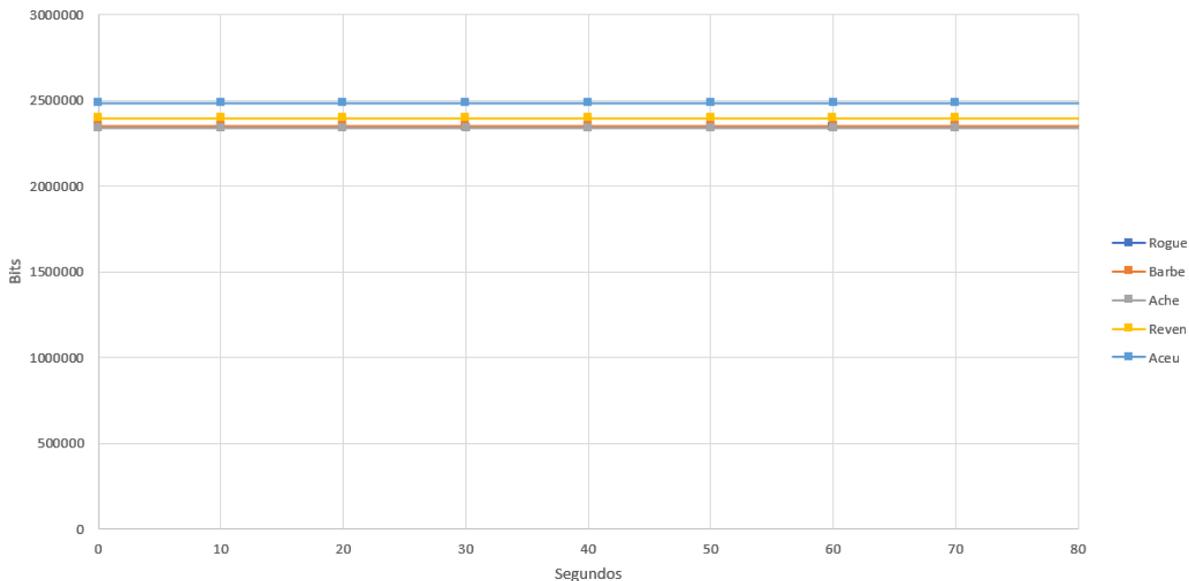


Fig. 11. Captura retransmisiones en la plataforma Twitch con calidad 720p

Por último, para la resolución de 480p (Fig.12) tenemos un rango mucho más concentrado de 1400-1600 kbps.

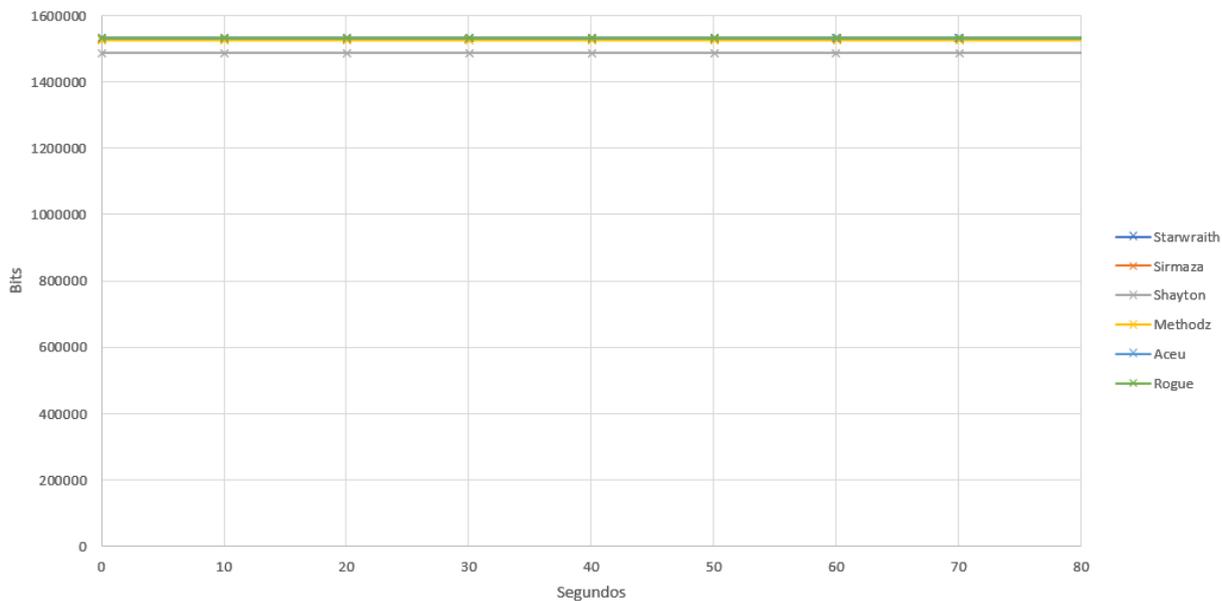


Fig. 12. Captura retransmisiones en la plataforma Twitch con calidad 480p

Los datos obtenidos de las capturas de tráfico de las graficas anteriores se resumen en la Tabla 4.

Resolución	Bitrate
1080p	4000-9000 kbps
720p60	3000-4000 kbps
720p30	2000-3000 kbps
480p	1400-1600 kbps

Tabla 4. parámetros reales de bitrate por calidad en Twitch.

VI.2.2. YOUTUBE

En este caso para la resolución de Full HD (Fig.13) sucede algo diferente que para la misma en Twitch, pues en este caso están más dispersos, aunque siguen entrando dentro de un rango similar. En la gráfica de Twitch (Fig.8) habíamos visto que se concentraban en dos franjas.

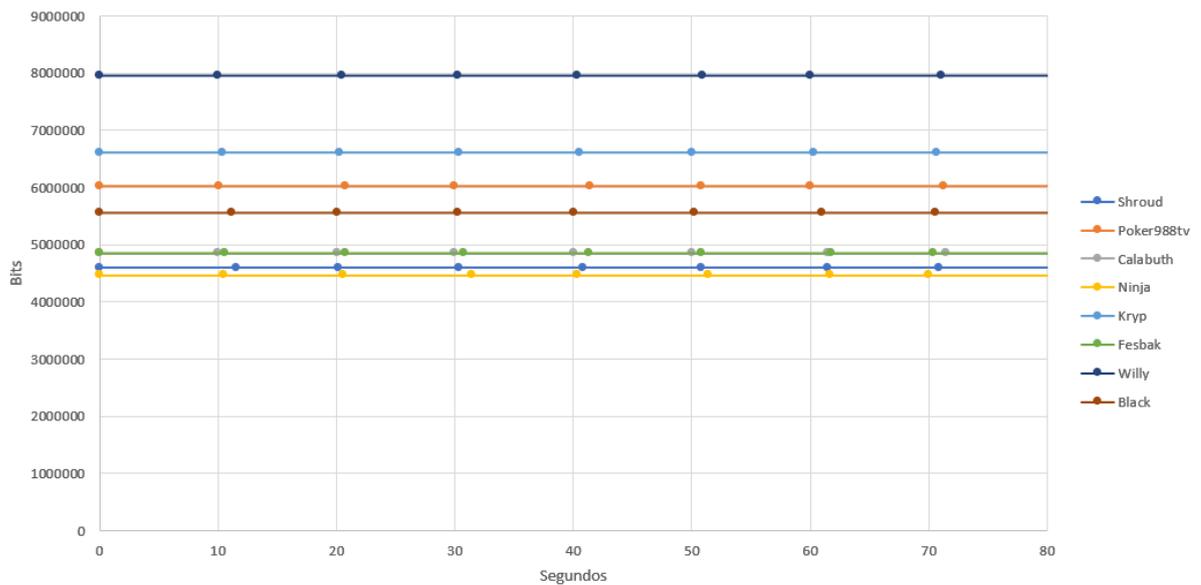


Fig. 13. Captura retransmisiones en la plataforma YouTube con calidad 1080p

Igual que sucedía para la resolución anterior, para 720p (Fig.14) también podemos ver que los datos son más dispersos. En Twitch habíamos dividido esta resolución en 2 pues estaban bien diferencias sin embargo aquí es más complicado y se ha decidido agrupar en una gráfica solo.

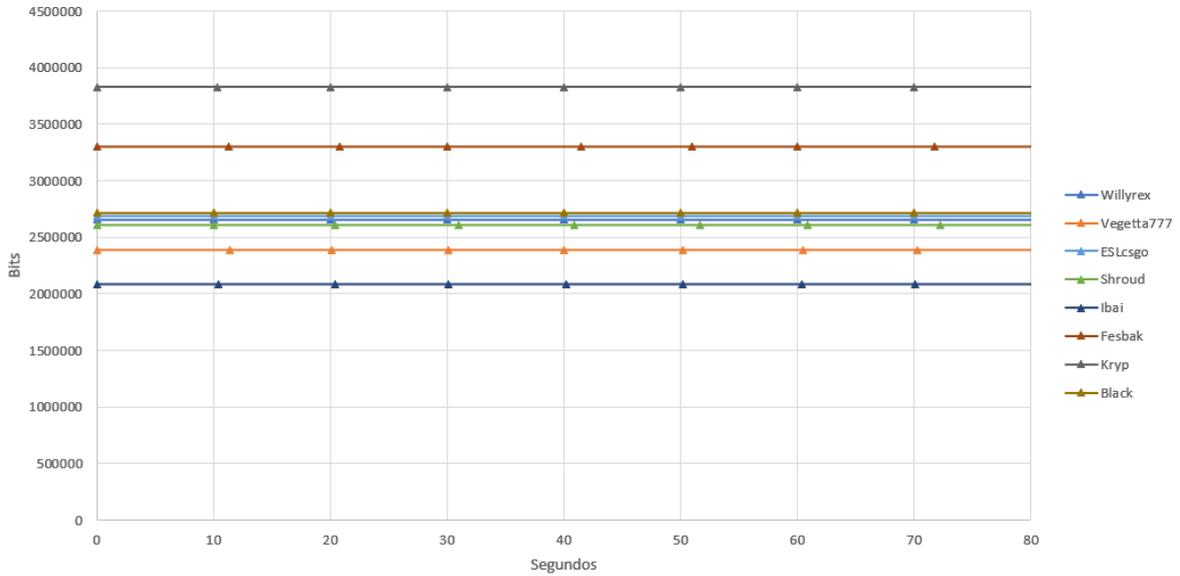


Fig. 14. Captura retransmisiones en la plataforma YouTube con calidad 720p

Sucede lo mismo en 480p (Fig.15) que para la resolución anterior. Se puede determinar que, en YouTube, los rangos son más espaciados, posiblemente sea debido a la variedad de formatos de compresión que utiliza.

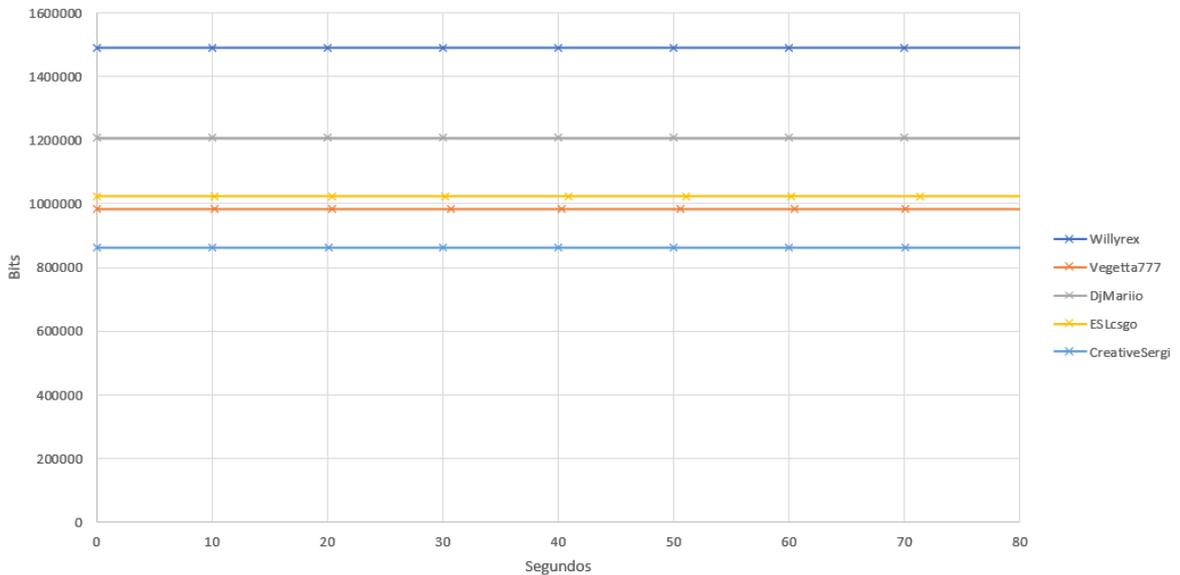


Fig. 15. Captura retransmisiones en la plataforma YouTube con calidad 480p

Los datos obtenidos de las capturas de tráfico de las gráficas de YouTube se resumen en la Tabla 5.

Resolución	Bitrate
1080p	4000-9000 kbps
720p	2250-4000 kbps
480p	800-1600 kbps

Tabla 5. parámetros reales de bitrate por calidad en YouTube.

VI.2.3. NETFLIX

Con Netflix han surgido diversos problemas. Hay contenido que divide su tráfico en varios puertos cercanos entre sí, pero es algo que no ocurre con todo el contenido y al dividirse ha originado que el rango sea más grande como se puede ver en la Fig.16.

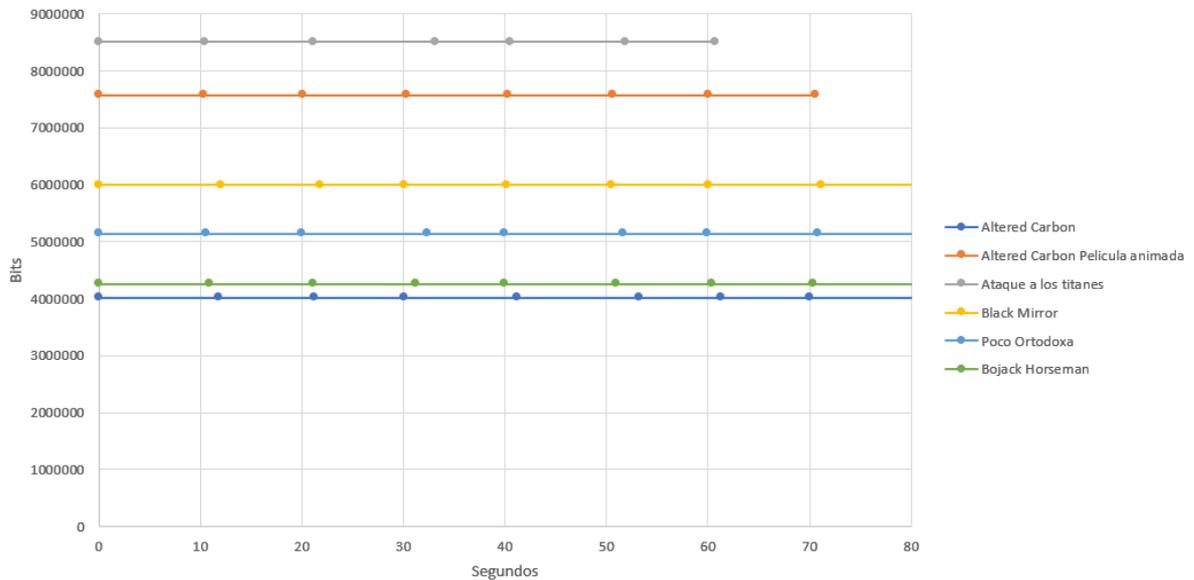


Fig. 16. Captura retransmisiones en la plataforma Netflix con calidad alta

VI.2.4. TV ONLINE

Se han visitado los diferentes servicios online gratuitos que ofrecen las cadenas televisión para visualizar su contenido desde sus respectivas páginas web. Este contenido, al ser gratuito tiene un límite de calidad de 720p, por tanto, será la calidad más utilizada entre los usuarios que frecuenten este tipo de páginas. Podemos ver su análisis en la Fig.17.

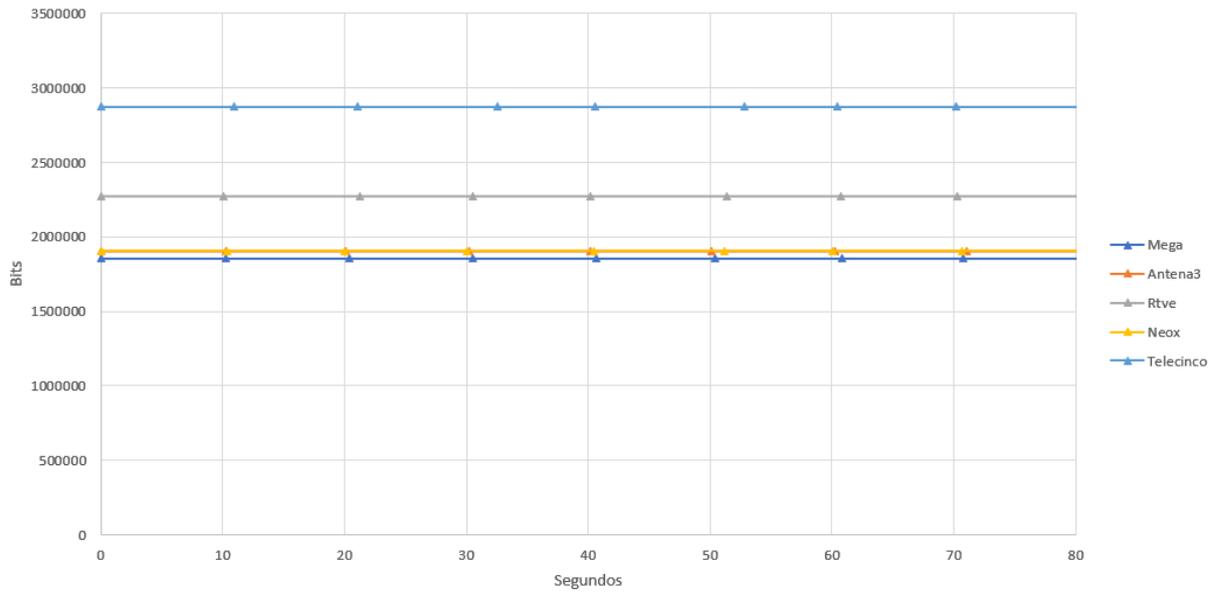


Fig. 17. Captura retransmisiones en plataformas de Tv online con calidad 720p

VI.2.5. FACEBOOK

En el caso de Facebook, vemos como ya los valores siguen dentro de los rangos que habíamos visto en anteriores plataformas como Twitch y YouTube para todas las resoluciones: 1080p (Fig.18), 720p (Fig.19), y 480p (Fig.20). Además, ya se empieza a visualizar un patrón.

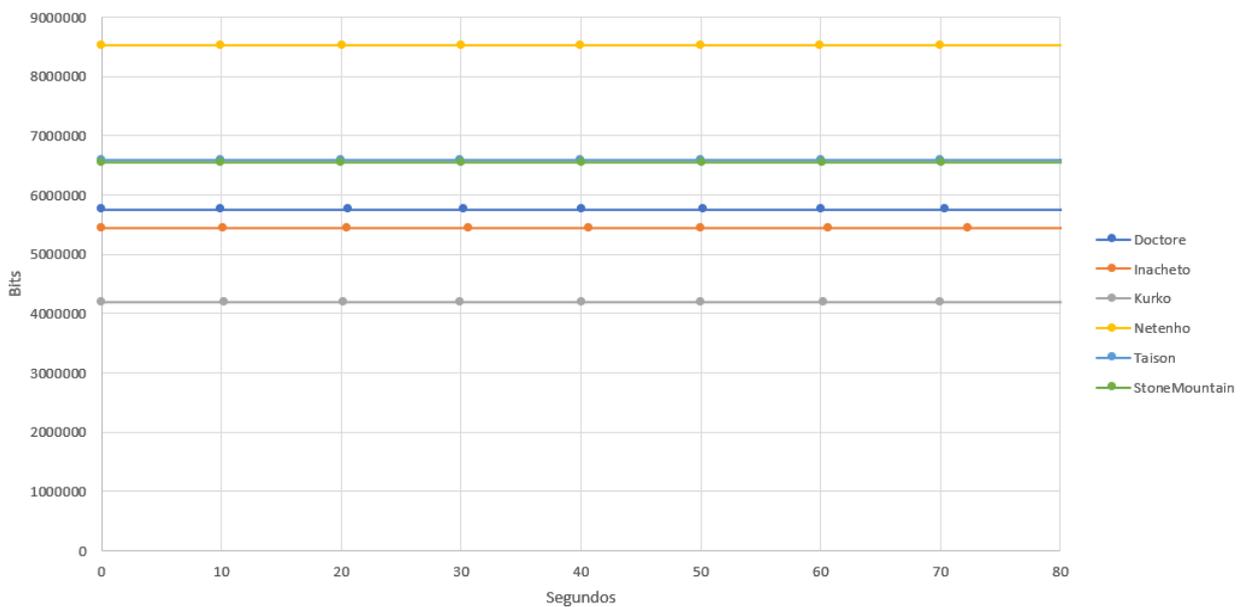


Fig. 18. Captura retransmisiones en la plataforma Facebook con calidad 1080p

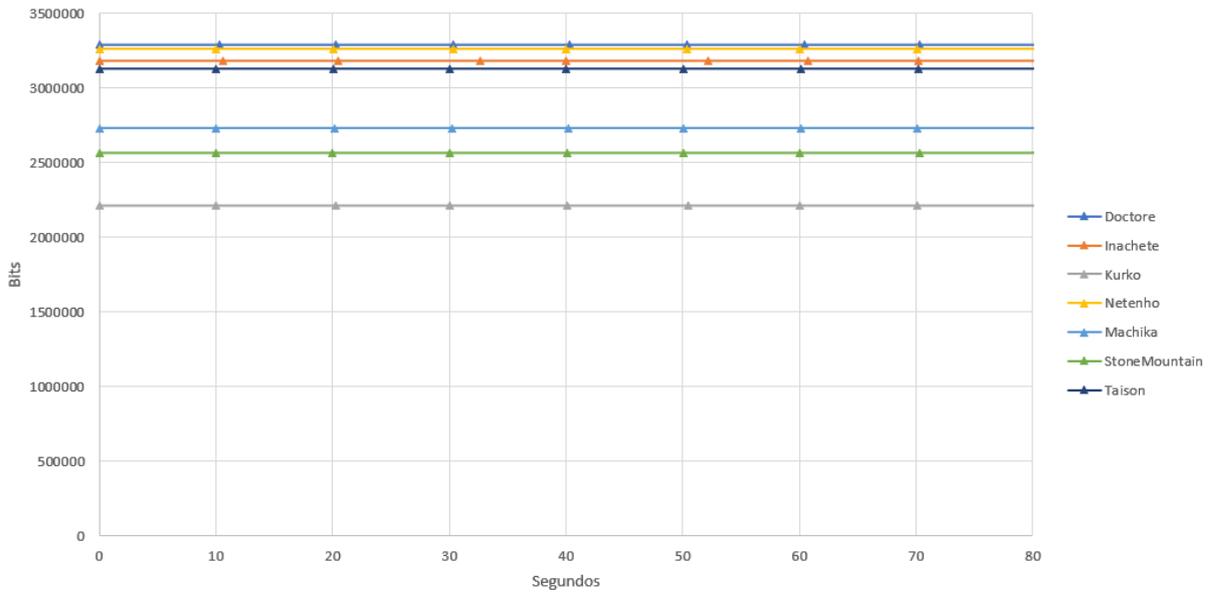


Fig. 19. Captura retransmisiones en la plataforma Facebook con calidad 720p

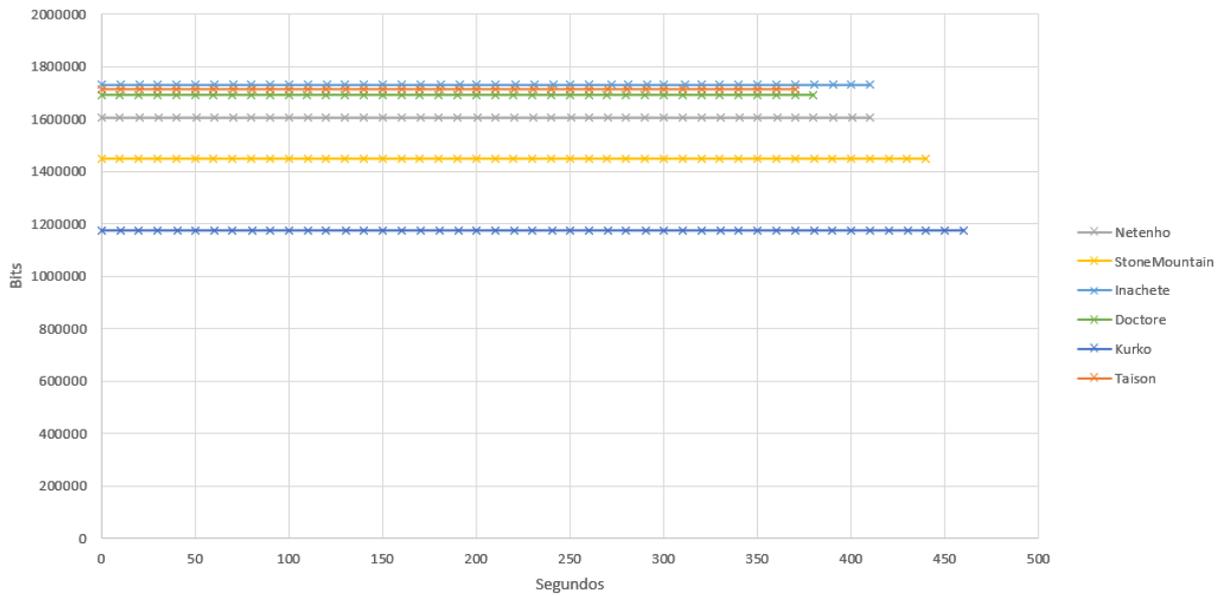


Fig. 20. Captura retransmisiones en la plataforma Facebook con calidad 480p

Los datos obtenidos de las capturas de Facebook se resumen en la Tabla 6.

Resolución	Bitrate
1080p	4000-9000 kbps
720p	2000-4000 kbps
480p	800-1800 kbps

Tabla 6. parámetros reales de bitrate por calidad en Facebook.

VI. CONCLUSIONES

VI.1 CUMPLIMIENTO DEL OBJETIVO

Una vez tenemos tanto los datos teóricos como los reales obtenidos mediante captura de tráfico, realizaremos las comparaciones.

Como se puede apreciar, los valores que nos dan en sus páginas web (apartado III.1) son muy ambiguos. Sin embargo, en los datos reales nos damos cuenta de que en la mayoría de los casos existe un rango específico para cada calidad que se repite sea cual sea la plataforma. Este rango es el patrón de comportamiento que estábamos buscando y es el que nos permitirá determinar un intervalo de bitrate para cada grado de calidad de imagen.

La tabla final con los resultados de los intervalos de bitrate asignados se puede observar en la Tabla 7.

Resolución	Bitrate
1080p	4000-9000 kbps
720p	1800-4000 kbps
480p	800-1800 kbps
< 480p	< 800 kbps

Tabla 7. Tabla final bitrate por calidad.

No se han analizado calidades inferiores a 480p pues cada vez son menos utilizadas, sin embargo, hay videos antiguos que tienen como máxima dicha calidad. Estos videos se detectarán como calidad inferior a 480p.

Una vez tenemos la tabla final, podremos identificar la cantidad de tráfico en resolución Full HD (1920x1080) o superior que se está consumiendo en una red o si, por el contrario, se tiene un consumo de HD (1280x720) o inferior.

Esta información nos podría dar, por ejemplo, qué porcentajes de calidad de video en consumo multimedia se suelen utilizar por diferentes regiones del mundo si analizáramos tráfico de cada una de ellas.

En el inicio de este documento, se informaba de que la principal meta era conseguir detectar cual es la calidad más probable a la que se está reproduciendo un video, y dado que se ha conseguido establecer el patrón mencionado para diferentes plataformas, podemos concluir que la meta ha sido alcanzada.

VI.2 CONCLUSIONES SOBRE EL PROYECTO

En este documento, hemos buscado similitudes en el video que se ofrece desde diferentes plataformas de transmisión de contenido multimedia, ya sea en vivo o bajo demanda. Para ello hemos tenido que capturar el tráfico de una gran cantidad de videos y representar los datos obtenidos para detectar algún patrón de comportamiento que pueda clasificar cuando un flujo de datos es potencialmente un video de una calidad determinada. Hemos descubierto que, para cada resolución, hay un intervalo de bitrate que se cumple en la mayoría de las situaciones, con diferentes tipos de contenido y desde diferentes plataformas OTT. De esta forma, con esta información podría ser más sencillo analizar, por ejemplo, para los proveedores de servicio, la resolución que utilizan la mayoría de los usuarios de determinadas zonas, información que podría ser beneficiosa para las ofrecer mejores servicios y hacer competencia, si vieran que en una zona se está utilizando una calidad de video baja por gran parte de los usuarios.

VI.3 PROBLEMAS ENCONTRADOS Y SOLUCIONES

Durante la realización del trabajo se han encontrado diversos problemas que han dificultado el correcto análisis de algunas plataformas.

Este es el caso de Netflix. Mientras estábamos analizando su tráfico, al contrario que sucedía en los demás servicios, este no concentra la mayor parte únicamente en un puerto, sino que se divide en varios puertos. Este incidente no ocurre en todos los videos sino en un pequeño porcentaje en comparación con los que si lo concentran en un único puerto.

Otro problema lo hemos encontrado en YouTube. Esta plataforma anunció en su blog para desarrolladores [27], que estaban trabajando en un nuevo método de compresión llamado VP9 el cual ofrece una calidad similar con un menor bitrate y que su integración se iría dando progresivamente hasta finalizar este 2020. Desde la propia plataforma podemos consultar el formato de compresión que está utilizando cada video, sin embargo, no hay un criterio oficial para utilizar una u otra, ya que la compresión AV1 también es utilizada en una gran cantidad de videos.

Esta incógnita en cuanto a la compresión puede generar grandes variaciones en la calidad del video, y es un hecho que han recibido numerosas quejas por parte de sus creadores de contenido, al ver que la calidad de sus videos, cuando son subidos a la plataforma, se reducía exageradamente.

Incluso con estos problemas, podemos ver que, aunque haya videos que se salgan de los parámetros establecidos, la gran mayoría sigue el mismo patrón.

VI.4 FUTURAS LINEAS DE TRABAJO

Una vez tenemos el patrón conseguido, como se ha mencionado durante el proceso de captura, este proceso se ha realizado mediante Ethernet. Para continuar con este proyecto, sería ideal realizar el mismo estudio desde diferentes dispositivos como smartphones, tablets, Smart Tv, etc para comprobar si afecta de alguna manera o si se mantiene en los rangos de bitrate establecidos en este trabajo.

Además, la posibilidad de utilizar estos datos en análisis de tráfico en tiempo real y en Big Data nos dará una vía más para catalogar la interacción de los usuarios con internet con mayor velocidad.

VII. REFERENCIAS

- [1] Página Web Sandvine. *The global Internet Phenome report. Octubre 2018*. Visto última vez el 29/11/220. Recuperado de: https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2018-phenomena-presentation-final.pdf
- [2] Página Web Sandvine. *The global Internet Phenome report. Septiembre 2019*. Visto última vez el 29/11/220. Recuperado de: https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/Internet%20Phenomena/Internet%20Phenomena%20Presentation%2020190910.pdf
- [3] Langley A, Riddoch A, Wilk A, Vicente A, Krasic C, Zhang D, Yang F, Kouranov F, Swett I, Iyengar J, Bailey J, Dorfman J, Roskind J, Kulik J, Westin P, Tenneti R, Shade R, Hamilton R, Vasiliev V, Chang WT, Shi Z (2017) *The QUIC transport protocol: design and internet-scale deployment*. In: Proceedings of the conference of the ACM special interest group on data communication. ACM, pp 183–196
- [4] Página web Datatracker IETF. Jana Iyengar , Martin Thomson: “*QUIC: A UDP-Based Multiplexed and Secure Transport*”. Octubre 2020. Visto última vez el 29/11/2020. Recuperado de: <https://datatracker.ietf.org/doc/draft-ietf-quic-transport/>
- [5] Blog de desarrolladores de Google. *A QUIC update on Google’s experimental transport. Abril 2015*. Visto última vez el 29/11/220. Recuperado de: <https://blog.chromium.org/2015/04/a-quic-update-on-googles-experimental.html>
- [6] Petr Matoušek, Jan Havlík, “*Video quality monitoring using netflow*”, Bachelor’s thesis, 2017.
- [7] T. Hoßfeld, R. Schatz, M. Seufert, M. Hirth, T. Zinner and P. Tran-Gia, “*Quantification of YouTube QoE via Crowdsourcing*”, International Workshop on Multimedia Quality of Experience-Modeling Evaluation and Directions (MQoE), 2011.
- [8] M. Seufert, T. Hoßfeld and C. Sieber, “*Impact of Intermediate Layer on Quality of Experience of HTTP Adaptive Streaming*”, 11th International Conference on Network and Service Management (CNSM), 2015.
- [9] S. Valenti, D. Rossi, A. Dainotti, A. Pescapè, A. Finamore and M. Mellia, “*Reviewing traffic classification in Data Traffic Monitoring and Analysis—From Measurement Classification and Anomaly Detection to Quality of Experience*, Heidelberg, Germany:Springer, 2013.
- [10] A. Ahmed, Z. Shafiq, H. Bedi and A. Khakpour, “*Suffering from buffering? Detecting QoE impairments in live video streams*,” 2017 IEEE 25th International Conference on Network Protocols (ICNP), Toronto, ON, 2017, pp. 1-10, doi: 10.1109/ICNP.2017.8117561.
- [11] S. Galetto et al., “*Detection of video/audio streaming packet flows for non-intrusive QoS/QoE monitoring*,” 2017 IEEE International Workshop on Measurement and Networking (M&N), Naples, 2017, pp. 1-6, doi: 10.1109/IWMN.2017.8078403.
- [12] Y. Chen, Q. Chen, F. Zhang, Q. Zhang, K. Wu, R. Huang, et al., “*Understanding viewer engagement of video service in Wi-Fi network*”, Computer Networks, vol. 91, pp. 101-116, 2015.

- [13] D. Joumblatt, J. Chandrashekar, B. Kveton, N. Taft and R. Teixeira, "Predicting User Dissatisfaction with Internet Application Performance at End-Hosts", IEEE INFOCOM, 2013.
- [14] F. Jackson, R. Amin, Y. Fu, J. E. Gilbert and J. Martin, "A User Study of Netflix Streaming", International Conference on Design User Experience and Usability (DUXU), 2015.
- [15] Antonio Lavado, "El consumo de YouTube en España", Global Media Journal México, Volumen 7, Número 14Pp. 76-92, 2010.
- [16] M. Seufert, P. Casas, N. Wehner, L. Gang and K. Li, "Stream-based Machine Learning for Real-time QoE Analysis of Encrypted Video Streaming Traffic," 2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN), Paris, France, 2019, pp. 76-81, doi: 10.1109/ICIN.2019.8685901.
- [17] Krishna Bhattar, Siddhi Gavhane, Priyanka Dhamne, Shardul Rabade, G. B. Aochar, "A Review on YouTube Data Analysis Using MapReduce on Hadoop", International Journal of Research in Engineering, Science and Management Volume-2, Issue-12, December-2019.
- [18] A. D'Alconzo, I. Drago, A. Morichetta, M. Mellia and P. Casas, "A Survey on Big Data for Network Traffic Monitoring and Analysis," in IEEE Transactions on Network and Service Management, vol. 16, no. 3, pp. 800-813, Sept. 2019, doi: 10.1109/TNSM.2019.2933358.
- [19] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin and K.-Y. Tung, "Intrusion detection system: A comprehensive review", J. Netw. Comput. Appl., vol. 36, no. 1, pp. 16-24, 2013.
- [20] T. T. T. Nguyen and G. Armitage, "A survey of techniques for Internet traffic classification using machine learning", IEEE Commun. Surveys Tuts., vol. 10, no. 4, pp. 56-76, 4th Quart. 2008.
- [21] A. Putina et al., "Telemetry-based stream-learning of BGP anomalies", Proc. Big-DAMA, pp. 15-20, 2018.
- [22] Página web YouTube. Ontologist. "VIDEO ROOM 1000 COMPLETE MIX -- All 1000 videos seen in sequential order!", Visto última vez 29/11/2020. Recuperado de: <https://www.youtube.com/watch?v=icruGcSsPp0>
- [23] Página web Streamlabs. *Streamlabs & Stream Hatchet Q2 2020 Live Streaming Industry Report*. Julio 2020. Visto última vez el 29/11/2020. Recuperado de: <https://blog.streamlabs.com/streamlabs-stream-hatchet-q2-2020-live-streaming-industry-report-44298e0d15bc>
- [24] Página web twitch. *Broadcasting Guidelines*. Última vez visto el 29/11/2020. Recuperado de: <https://stream.twitch.tv/encoding/>
- [25] Página web Youtube. *Elegir la configuración del codificador, tasas de bits y resoluciones para emisiones en directo*. Visto última vez el 29/11/2020. Recuperado de: <https://support.google.com/youtube/answer/2853702?hl=es>

- [26] Página web Facebook. *Información sobre el streaming en HD*. Última vez visto 29/11/2020. Recuperado de: <https://www.facebook.com/business/help/440620969777882?id=648321075955172>
- [27] Blog de desarrolladores de Youtube. *VP9: Faster, better, buffer-free YouTube videos*. Abril 2015. Visto última vez el 29/11/2020. Recuperado de: <https://youtube-eng.googleblog.com/2015/04/vp9-faster-better-buffer-free-youtube.html>