(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 995 015**

(51) Int. Cl.:

H04L 43/026 (2012.01)
H04L 43/028 (2012.01)
H04L 43/0817 (2012.01)
H04L 43/12 (2012.01)
H04L 43/16 (2012.01)
H04L 43/20 (2012.01)
H04L 12/40 (2006.01)
H04L 12/46 (2006.01)
H04L 43/045 (2012.01)
H04L 43/065 (2012.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **21.09.2018 PCT/AU2018/051036**

(87) Fecha y número de publicación internacional: **04.04.2019 WO19060949**

(96) Fecha de presentación y número de la solicitud europea: **21.09.2018 E 18860240 (3)**

(97) Fecha y número de publicación de la concesión europea: **06.11.2024 EP 3688944**

(54) Título: **Procedimiento, medio de almacenamiento legible por ordenador y sistema para identificar y clasificar datos de vídeo**

(30) Prioridad:

27.09.2017 AU 2017903915

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
05.02.2025

(73) Titular/es:

CANOPUS NETWORKS ASSETS PTY LTD (100.00%)
Level 5, 24-28 Campbell Street
Haymarket, New South Wales 2000, AU

(72) Inventor/es:

SIVARAMAN, VIJAY;
GHARAKHEILI, HASSAN HABIBI y
WANG, YU

(74) Agente/Representante:

ISERN JARA, Jorge

DESCRIPCIÓN

Procedimiento, medio de almacenamiento legible por ordenador y sistema para identificar y clasificar datos de vídeo

Campo técnico

La presente invención se refiere a procedimientos y un sistema para identificar y clasificar datos de red y, particularmente, pero no exclusivamente, a un procedimiento y sistema para identificar y clasificar datos de video, datos de realidad aumentada, datos de realidad virtual y otros flujos de datos grandes que se desplazan a través de una red.

Antecedentes

La tecnología de redes de Internet ha revolucionado nuestras vidas en las últimas décadas. Un proveedor de la red de Internet proporciona a sus usuarios la capacidad de acceder al contenido de diversas fuentes, y el contenido descargado por los usuarios típicamente incluye datos de audio, datos de video, mensajería en línea, navegación en sitios web, navegación en redes sociales y transferencia de archivos (por ejemplo, que incluye el uso de Facebook™, Instagram™, y WhatsApp™) y así sucesivamente.

Es conveniente que los proveedores de la red comprendan cómo se usa su red y qué tipo de contenido acceden los usuarios. Las grandes transmisiones de datos, principalmente de datos de video, constituyen la mayor parte del tráfico de la red hoy en día. Actualmente, los proveedores de la red tienen una visibilidad muy limitada del tráfico que se desplaza a través de su red, y esta visibilidad limitada obstaculiza la capacidad del proveedor de la red para identificar y resolver los problemas de capacidad de datos que enfrenta la red. Actualmente, abordan los problemas de capacidad de datos al aumentar el ancho de banda de sus redes, lo cual es una solución costosa.

Para gestionar mejor el tráfico de datos (por razones de calidad y costo), sería ventajoso para los proveedores de la red tener visibilidad sobre aspectos microscópicos, tales como cuántos flujos de video están activos simultáneamente en un momento dado, cuáles son sus duraciones, en qué resoluciones operan y con qué frecuencia adaptan su velocidad. La visibilidad de estos atributos puede permitirles comprender mejor tanto las características del contenido como los patrones de visualización de datos, por lo que pueden implementar cambios útiles para ajustar su red para cumplir con las expectativas del proveedor de contenido y mejorar la experiencia del usuario.

Actualmente, se están utilizando dos tecnologías principales para comprender el tráfico de la red. La primera tecnología es basada en hardware y se conoce como Inspección Profunda de Paquetes (DPI). Esta tecnología analiza cada uno y cada paquete que se desplaza a través de la red mediante el uso de hardware que es muy costoso, tanto económica como computacionalmente (debido a que se requiere una alta potencia de procesamiento para analizar cada uno y cada paquete). Otra desventaja de esta técnica es que proporciona una escalabilidad limitada. Por al menos estas razones, DPI no es práctico para implementar para la mayoría de los operadores de la red.

La segunda tecnología hace uso del software de inspección de paquetes para el análisis de paquetes y el análisis de flujo por separado. Esta técnica también es computacionalmente muy costosa y tiene una escalabilidad limitada.

El tráfico de vídeo está aumentando rápidamente cada día y se supone que aumentará aún más en el futuro cercano a medida que se logren resoluciones más altas (por ejemplo, 1.440 p y 4 K) se vuelven más prevalentes, y la realidad aumentada y virtual comienzan a despegar.

Por al menos las razones anteriores, los proveedores de la red necesitan una mejor visibilidad de sus redes, en particular para resolver problemas de capacidad de la red de una manera eficiente y rentable, y para mejorar la experiencia del usuario.

El tema relevante para la presente invención se describe en el documento US 2015/071072 A1, y en el documento publicado de Uddin Mostafa y otros, "Trafficvision: A Case for Pushing Software Defined Networks to Wireless Edges", 2016 IEEE 13TH International conference on mobile ad hoc and sensor systems (MASS), XP033042640.

Por lo tanto, se desea proporcionar un procedimiento y sistema de monitoreo del tráfico de red que supere o alivie una o más dificultades de la técnica anterior, o al menos proporcionar una alternativa útil.

Sumario

De acuerdo con una realización de la presente invención, se proporciona un procedimiento de monitoreo del tráfico de red que se ejecuta por un sistema de monitoreo del tráfico de red de una red de comunicaciones, de acuerdo con la reivindicación 1. En algunas realizaciones, los tipos de flujo incluyen flujos de video y flujos que no son de video.

En algunas realizaciones, los tipos de flujo incluyen flujos de video de respectivas resoluciones diferentes.

En algunas realizaciones, el procedimiento incluye determinar los proveedores de servicios de al menos algunos de los flujos grandes de red a partir de la información de DNS.

En algunas realizaciones, las métricas de flujo incluyen métricas de la ráfaga en las respectivas escalas de tiempo.

En algunas realizaciones, las escalas de tiempo representan una serie geométrica.

De acuerdo con una realización de la presente invención, se proporciona un procedimiento de monitoreo del tráfico de red que se ejecuta por un conmutador de flujo de red definida por software (SDN) de una red de comunicaciones, de acuerdo con la reivindicación 7.

De acuerdo con algunas realizaciones de la presente invención, se proporciona un sistema de monitoreo del tráfico de la red de acuerdo con la reivindicación 9.

En algunas realizaciones, el sistema incluye un componente de interfaz de usuario configurado para recibir las solicitudes de los usuarios y, en respuesta a las solicitudes, generar los datos de la interfaz de usuario que representan una interfaz de usuario interactiva para mostrar la información en los flujos grandes de red detectados por el sistema, la información que incluye las clasificaciones de los flujos grandes de red.

De acuerdo con algunas realizaciones de la presente invención, se proporciona al menos un medio de almacenamiento legible por ordenador que tiene almacenadas en el mismo las instrucciones ejecutables, de acuerdo con la reivindicación 8.

Se conoce que el procedimiento convencional de inspección profunda de paquetes monitorea cada paquete individual de cada flujo de datos. Teniendo en cuenta el tamaño del tráfico de la red, esto no es escalable, lleva mucho tiempo y es extremadamente costoso de implementar.

También se describe en la presente memoria, el monitoreo del tráfico de la red se logra al combinar el monitoreo a nivel de paquetes con el monitoreo a nivel de flujo. En una realización, la etapa de monitorear los datos de un flujo de datos comprende la etapa de obtener paquetes de datos hasta que se alcance un umbral. Ventajosamente, en una realización, por lo tanto solo algunos de los paquetes de datos se monitorean (por ejemplo, los primeros MegaBytes de cada flujo de datos). En una realización, esta inspección limitada de paquetes proporciona suficiente información para determinar el tipo de datos, la información del proveedor de contenido, la dirección de la solicitud de contenido, y así sucesivamente. En una realización, esto se sigue por el monitoreo a nivel de flujo de la secuencia de datos, que puede usarse para implementar un análisis de clasificación para clasificar las secuencias de datos del tipo de datos identificado en diferentes categorías de datos.

En una realización, el umbral se elige para activar el monitoreo del nivel de flujo para los tipos de datos que comprenden flujos de datos de gran volumen. Estos se conocen de cualquier otra manera como "elefantes", e incluyen descargas grandes, transmisión de video, realidad aumentada, flujos de datos de realidad virtual y otros flujos de datos grandes. Los flujos de datos que no alcanzan el umbral generalmente comprenden pequeños flujos de datos ("ratones"), tales como publicaciones de redes sociales y similares. Mientras que los ratones comprenden la mayoría de los tipos de datos, los elefantes ocupan la mayor parte del volumen de tráfico de datos. Ignorando los ratones, significa que es posible que las realizaciones de la invención se concentren en los elefantes, que son los flujos grandes de datos.

El procedimiento descrito en la presente memoria es altamente escalable porque solo un número limitado de paquetes de datos de un flujo de datos se someten a monitoreo a nivel de paquete. Esto proporciona una solución de bajo costo y altamente escalable para monitorear y clasificar el tráfico de la red. Además, el procedimiento se concentra en los flujos de datos de gran volumen e ignora los ratones, optimizando además el procesamiento.

La etapa de obtener los paquetes de datos puede comprender replicar los paquetes de datos del flujo de datos.

Es una ventaja de al menos una realización de la presente invención que los paquetes de datos del flujo de datos que se examina no se vean afectados o modificados. Esto se debe a que la inspección de paquetes se realiza en paquetes de datos replicados.

Como se describe en la presente memoria, la etapa de obtener los paquetes de datos de un flujo de datos se detiene cuando se alcanza el umbral y se determina el tipo de datos del flujo de datos.

Como se describe en la presente memoria, la etapa de monitorear los datos se implementa a través de una solución de red definida por software (SDN). Como se describe en la presente memoria, la telemetría de flujo se implementa utilizando contadores de hardware.

El equilibrio entre el hardware y el procesamiento de software reduce los costos, aumenta la escalabilidad y permite la extracción de suficiente información de los datos para la implementación de un análisis de clasificación.

Como se describe en la presente memoria, el procedimiento comprende la etapa adicional de llevar a cabo un análisis de clasificación para clasificar el tipo de datos predeterminado en una de una pluralidad de categorías de datos. Las categorías de datos pueden comprender la clasificación en la resolución de datos, por ejemplo, alta definición, definición media, definición baja. Las categorías también pueden comprender datos relacionados con una identidad del proveedor (por ejemplo, Netflix™, YouTube™ etc.). La categorización puede comprender la identificación del tipo de datos, por ejemplo, video, descarga grande, etc.

Como se describe en la presente memoria, la clasificación del tipo de datos predeterminado se basa en las características de los datos, que comprende uno o más de: perfil escaneado, tamaño de flujo de datos, resolución e información del proveedor de datos.

Como se describe en la presente memoria, se implementa un procedimiento de aprendizaje automático para mejorar el análisis de clasificación.

También se describe en la presente memoria un aparato para monitorear el tráfico de datos en una red, el tráfico de datos que comprende una pluralidad de flujos de datos, el aparato que comprende al menos un procesador dispuesto para monitorear los datos en cada flujo de datos y determinar el tipo de datos para cada flujo de datos, y dispuesto para implementar la telemetría de flujo para al menos uno de los tipos de datos predeterminados, para determinar el volumen de flujo para cada flujo de datos del tipo de datos predeterminado.

El procesador puede comprender una aplicación de red definida por software que se dispone para instruir la obtención de paquetes de datos del flujo de datos hasta que se alcance un umbral. En una realización, el procesador comprende un detector de flujo grande dispuesto para inspeccionar los paquetes de datos. En una realización, los paquetes de datos se obtienen al replicar los paquetes de datos del flujo de datos.

El procesador se dispone para recopilar contadores de hardware para implementar la telemetría de flujo.

Como se describe en la presente memoria, el procesador se dispone para examinar el tipo de datos predeterminado y clasificar las transmisiones de datos del tipo de datos en una de una pluralidad de categorías de datos. En una realización, el aprendizaje automático se usa para perfilar las secuencias de datos y categorizarlas. Se proporciona una base de datos para almacenar las características de las transmisiones de datos para permitir la clasificación.

En una realización, el aparato comprende una interfaz de usuario que presenta información sobre los tipos de datos y categorías y el análisis del flujo de datos.

También se describe en la presente memoria un programa informático, dispuesto para instruir a un procesador para implementar cualquiera de los procedimientos anteriores.

En la presente memoria se describe además un medio legible por ordenador no volátil, que proporciona un programa informático de acuerdo con el tercer aspecto de la invención.

En la presente memoria se describe además una señal de datos que comprende cualquiera de los programas informáticos anteriores.

Breve descripción de las figuras

Algunas realizaciones de la presente invención se describen a continuación, a manera de ejemplo solamente, con referencia a las figuras adjuntas, en las cuales:

La Figura 1 muestra la arquitectura y los bloques funcionales de un sistema de monitoreo del tráfico de red de acuerdo con las realizaciones descritas de la presente invención;

La Figura 2 muestra la estructura de la tabla de flujo de un conmutador SDN del sistema, de acuerdo con una realización;

La Figura 3 muestra una comparación entre varios perfiles de tráfico observados para varios flujos de video proporcionados por diferentes proveedores de video;

La Figura 4 muestra la arquitectura y los bloques funcionales de un sistema de monitoreo del tráfico de red de acuerdo con una realización de la presente invención;

Las Figuras 5 y 6 muestran instantáneas de una interfaz web proporcionada a los administradores de la red para visualizar la información relacionada con las transmisiones de vídeo en su red;

La Figura 7 muestra histogramas de tiempo inactivo, velocidad promedio y variabilidad en varias escalas de tiempo para flujos de video frente a no video;

La Figura 8 muestra un histograma de tiempo inactivo, velocidad promedio y variabilidad en varias escalas de tiempo para varias resoluciones de flujos de video;

La Figura 9 muestra una matriz de confusión de un identificador de vídeo del sistema;

La Figura 10 muestra una matriz de confusión de un clasificador de resolución de video del sistema;

La Figura 11 muestra la precisión del rendimiento del sistema. La Figura 11(a) muestra el mérito de los atributos, la Figura 11(b) muestra la precisión de la identificación de vídeo y la Figura 11(c) muestra la precisión de la clasificación de resolución;

La Figura 12 muestra una configuración usada para la evaluación del rendimiento del sistema;

La Figura 13 muestra la carga de la red para varios proveedores de contenido a intervalos de un segundo;

La Figura 14 muestra estadísticas de flujo que indican la detección de flujos de elefante por el sistema;

La Figura 15 muestra la distribución del Consumo de Vídeo de Residencia (para el mes de mayo de 2017), que muestra (a) un gráfico circular de la fracción de transmisiones totales de los proveedores de video populares, (b) un gráfico de barras que muestra el número diario de transmisiones de video y (c) un gráfico de barras del número de resolución de video por hora;

La Figura 16 muestra la CCDF de las características de Vídeo de Dorm;

La Figura 17 es un diagrama de flujo de un procedimiento de detección de flujo grande del sistema; y

La Figura 18 es un diagrama de flujo de un procedimiento de clasificación de vídeo del sistema.

Descripción detallada

Las realizaciones de la presente invención incluyen un sistema y procedimiento de monitoreo del tráfico de red que son capaces de clasificar los paquetes de datos que fluyen a través de una red de comunicaciones en diferentes flujos de red, y de caracterizar esos flujos por tipo y propiedades del tráfico. Aunque algunas realizaciones de la presente invención se describen a continuación en el contexto de monitorear flujos de datos de video en una red de comunicaciones, debe entenderse que el aparato y el procedimiento de monitoreo del tráfico de red no se limitan a datos de video, sino que pueden aplicarse de manera general para identificar y caracterizar flujos de cualquier tipo de tráfico de red en una red de comunicaciones.

La Red Definida por Software (SDN) es una tecnología de red flexible y versátil que usa un sistema de control centralizado que se separa de los conmutadores de red y otros dispositivos de red. El sistema de control de SDN centralizado usa un protocolo de control de SDN tal como OpenFlow para configurar los dispositivos de la red SDN tales como los conmutadores de red. En la red convencional, cada conmutador tiene su propio software de control independiente para decidir dónde mover los paquetes de datos. Sin embargo, en un sistema SDN, las decisiones del movimiento de los paquetes se toman en última instancia por el controlador SDN centralizado que controla el comportamiento de los conmutadores SDN para procesar los paquetes en consecuencia. El controlador SDN se puede programar de forma personalizada, en base a las necesidades del operador de la red e independientemente de los conmutadores individuales.

Un conmutador SDN generalmente incluye tablas de flujo que definen reglas de coincidencia para identificar si un paquete de red recibido en un puerto de entrada del conmutador pertenece a cualquiera de una pluralidad de flujos definidos o predeterminados (también conocidos en la técnica como 'flujos de paquetes', 'flujos de red' y 'flujos de tráfico'), y para cada flujo, una acción para realizar en los paquetes que pertenecen al flujo, identificando típicamente un puerto de salida correspondiente del conmutador al que se deben emitir los paquetes de ese flujo desde el conmutador. Como se indicó anteriormente, las tablas de flujo de un conmutador SDN pueden modificarse dinámicamente por un controlador SDN a través de un protocolo de control SDN tal como el protocolo Open Flow.

Los inventores han determinado que un sistema basado en SDN es adecuado para identificar y clasificar los flujos de tráfico de red (que incluyen los flujos de tráfico de vídeo) que atraviesan una red de comunicaciones. Los inventores han desarrollado un aparato basado en SDN que incluye un controlador programable de forma independiente y conmutadores SDN, que en las realizaciones descritas son conmutadores OpenFlow de bajo costo disponibles en el mercado. Este sistema opera a una velocidad mucho mayor en comparación con los procedimientos convencionales de DPI y de software de inspección de paquetes.

1. Diseño y arquitectura del sistema

La Figura 1(a) muestra la arquitectura y los bloques funcionales de un aparato de monitoreo del tráfico de red aplicado a una red portadora, de acuerdo con una realización de la presente invención. En esta realización, el aparato de monitoreo del tráfico de red se puede insertar de manera transparente entre dos puertos de una red donde se desea el monitoreo del tráfico de red (monitoreo de video en la realización descrita). El aparato 20 se inserta entre una puerta de enlace de Internet 21 y una puerta de enlace de acceso 22 de la red. El usuario final (a la izquierda de la Figura 1(a)) se puede conectar a la red a través de la puerta de enlace de acceso 22 mediante el uso de tecnología alámbrica (DSL, Ethernet, Fibra) y/o inalámbrica (por ejemplo, 3G/4G, WiFi). Los proveedores de contenido de vídeo están a la derecha, conectados a la red de la empresa/portadora a través de la puerta de enlace de Internet 21.

El aparato 20 se puede insertar en cualquier enlace deseado como un 'bache en el cable' donde se requiere la inspección de datos de la red.

Como se muestra en la Figura 1(a), el aparato 20 incluye un conmutador SDN 23, un detector de flujo grande 24, un intermediario de datos 25, una interfaz de usuario 26, una base de datos 27 y una aplicación SDN 28 en un controlador SDN 29.

El tráfico de red del proveedor de contenido entra en el aparato 20 desde la puerta de enlace de Internet 21, y sale en la puerta de enlace de acceso 22 y hacia el usuario final. Típicamente, el tráfico de la red incluye todo tipo de flujos de datos, incluidos los archivos de video transmitidos en directo, los archivos de audio transmitidos en directo, los archivos de descarga grandes, los flujos de datos pequeños que representan la navegación en redes sociales y la mensajería de aplicaciones móviles, etc.

En la realización descrita, los archivos de video transmitidos por los usuarios a través de la red se monitorean de la siguiente manera.

En una instalación de campo nuevo de ejemplo, el conmutador SDN 23 se configura inicialmente para replicar todos los paquetes de datos de cada flujo entrante al detector de flujo grande 24. El detector de flujo grande 24 realiza un seguimiento del volumen de cada flujo hasta que se alcanza o excede un volumen de flujo umbral predeterminado, y luego notifica al intermediario de datos. En una realización, el volumen umbral predeterminado está en el rango de 2 a 20 Mega-bytes, en función del tipo de flujo de vídeo que se va a identificar. En otra realización, el volumen umbral se establece en 4 MegaBytes. Si el volumen de flujo es mayor que el umbral correspondiente, entonces se considera que es un "flujo pesado" (o como un "flujo de elefante", usando un término técnico). El flujo pesado puede ser un flujo de video o un archivo descargable de gran tamaño o un video descargable cuyo volumen de flujo y duración son mayores que el umbral predeterminado de volumen y período. Una vez que se identifica un flujo de elefante, el corredor de datos 25 instruye a la aplicación SDN 28 para insertar una entrada de flujo reactivo para este flujo específico en el conmutador SDN 23, y para detener el espejeo de paquetes para este flujo. Esto alivia al detector de flujo grande 24 de realizar un análisis adicional del flujo de elefante. Como un resultado, la escalabilidad del detector de flujo grande 24 se mejora sustancialmente en comparación con los sistemas convencionales de DPI y de inspección de software.

Una vez que se ha identificado un flujo de elefante y se guarda una entrada reactiva para el flujo de elefante en una tabla de flujo de la conmutación SDN 23, el corredor de datos 25 consulta los contadores de la conmutación SDN 23 periódicamente para desarrollar un perfil de tráfico para este flujo de elefante. En esta memoria descriptiva, un perfil de tráfico de un flujo incluye información sobre la identidad del flujo y la identidad del proveedor de contenido de ese flujo. Las Figuras 1(b) y 1(c) representan respectivamente los módulos internos del intermediario de datos 25 y la Aplicación SDN 28 que recopilan telemetría, desarrollan perfiles de tráfico y realizan los procedimientos de identificación y clasificación de flujo. El intermediario de datos 25 incluye dos procedimientos inteligentes, específicamente: (i) un identificador de vídeo y (ii) un clasificador de vídeo. Diferentes tipos de flujos de elefante tienen diferentes perfiles de velocidad de tráfico. En base a estas características, el identificador de vídeo se usa para identificar las secuencias de vídeo de los otros tipos de flujos de tráfico de los flujos de elefante identificados. Además, el clasificador de video se usa para clasificar las transmisiones de video identificadas por sus resoluciones.

El conmutador SDN 23 se comunica con el controlador SDN 29 mediante el uso de un protocolo OpenFlow. El conmutador SDN 23 actúa como un filtro de hardware que limita la fracción del tráfico (típicamente a los primeros MegaBytes del tráfico de un flujo) replicado para el análisis del flujo, mientras que la aplicación SDN 28 crea entradas de tabla de flujo reactivas para los flujos de elefante que luego se monitorean a través de los contadores de hardware y la Tabla 3 (Grupo). Los umbrales se ajustan en el volumen y la duración del flujo en el que se crea una entrada de flujo reactivo, y los inventores han descubierto empíricamente que un valor de 4 MegaBytes para el umbral de volumen funciona bien; esto mantiene las operaciones de modificación del flujo de hardware en menos del 1 % de todos los flujos (en los ensayos de los inventores, más del 99 % de los flujos son cortos), mientras que limita el reflejo de paquetes al detector de flujo grande 24 a menos de un tercio del tráfico del enlace (ya que alrededor del 75 % del volumen de tráfico se transporta en los flujos de elefante). Este equilibrio entre el hardware y el procesamiento de software reduce el costo, aumenta la escalabilidad y permite la extracción de suficiente

información para que los algoritmos de aprendizaje automático logren una alta precisión de clasificación.

2. Gestión de la tabla de flujo

La Figura 2 ilustra una estructura de tabla de flujo múltiple del conmutador SDN 23. Estas tablas de flujo de la conmutación SDN 23 se configuran para identificar y categorizar los flujos entrantes. La Tabla 0 y la Tabla 1 son una tabla de flujo reactivo y una tabla de flujo proactivo, respectivamente, y se usan para almacenar entradas reactivas y proactivas, respectivamente. La tabla 2 es una tabla de flujo predeterminada y la tabla 3 es una tabla de grupo. Mediante el uso de las tablas de flujo, se usa un comando de coincidencia para identificar los flujos entrantes conocidos, y los comandos de acción correspondientes se usan para realizar una acción apropiada de mover el flujo a la entrada correspondiente en la tabla de grupo (Tabla 3).

Las reglas reactivas de la Tabla 0 coinciden con 5 tuplas para flujos conocidos. Una 5-tupla es un conjunto ordenado de cinco valores que identifican un flujo. Las reglas reactivas de la Tabla 0 tienen la máxima prioridad y se instalan como consecuencia de los flujos de elefante identificados por el detector de flujo grande 22. Se desconectan automáticamente (y se eliminan de la tabla) tras un período de inactividad predefinido que varía de 10 segundos a 60 segundos. Las entradas de flujo reactivo logran dos objetivos: (i) detener la réplica de paquetes de flujo de elefante al detector de flujo grande de software 24, y (ii) proporcionar telemetría de nivel de flujo (características de flujo) para los flujos de elefante individuales (potencialmente de vídeo). La acción correspondiente a una coincidencia en la tabla reactiva (Tabla 0) envía el flujo a su entrada apropiada en la tabla de grupo (Tabla 3), que identifica el proveedor de contenido (YouTube, Netflix, etc.). El proveedor de contenido para el flujo se identifica mediante la búsqueda de la dirección IP del servidor en los sufijos DNS más recientes capturados (por ejemplo, googlevideo.com o nflxvideo.com) que se almacenan en una tabla de base de datos de series temporales (la "base de datos de flujo" en la Figura 1(b)) por el detector de flujo grande 24. Si se detecta un flujo de video de un nuevo sufijo DNS (por ejemplo, ttnvw.net), entonces se crea una nueva entrada de grupo (para Twitch en este ejemplo) dinámicamente en la tabla de grupos. Esto no solo hace que el aparato 20 sea adaptable a nuevos proveedores de contenido de vídeo, sino que también permite el seguimiento de los volúmenes de vídeo agregados para cada proveedor de contenido de vídeo al almacenarlos en la tabla de grupo. Por lo tanto, la tabla de flujo reactivo se usa para la visibilidad de grano fino, mientras que la tabla de grupo se usa para la visibilidad de nivel grueso de los flujos de video detectados por el aparato 20.

Las entradas proactivas (Tabla 1) se empujan de forma estática por el controlador SDN 29 de modo que todos los paquetes de Protocolo de Control de Transmisión (TCP) (proto=6) y Protocolo de Datagramas de Usuario (UDP) (proto=17) recibidos del proveedor de contenido, que ya no coinciden con un flujo de elefante (Tabla 0), se reenvían al puerto-2 (es decir, la puerta de enlace de acceso 22) y se replican en el puerto-3 por el conmutador SDN 23 al detector de flujo grande 24. Esto incluye paquetes de respuesta DNS que contienen los nombres de dominio de los proveedores de contenido de video y las direcciones IP del servidor de video. Todos los demás tipos de paquetes se envían a la Tabla 2, donde la acción predeterminada es conectar de manera cruzada los puertos de entrada (puerta de enlace de Internet 21) y salida (puerta de enlace de acceso 22) sin realizar ningún espejado o procesamiento.

El aparato 20 no envía ningún paquete de datos al controlador SDN 29, minimizando así la carga en el controlador SDN 29, reduciendo la latencia de reenvío de paquetes e inmunizando contra fallos del controlador SDN 29.

Es una ventaja del aparato 20 que es completamente transparente a la red. Esto se debe a que el conmutador de SDN 23 hace copias de los paquetes que requieren monitoreo y los envía al detector de flujo grande 24. El conmutador SDN 23 reenvía una copia de los paquetes de datos a su ruta de tráfico sin interrupción. El aparato 20 no modifica los paquetes.

Otra ventaja del aparato 20 es que no sobrecarga el controlador de SDN 29. El conmutador SDN 23 no envía ningún paquete de datos al controlador SDN 29; en cambio, cualquier paquete que necesite inspeccionarse se envía como copia al detector de flujo grande 24. Esto protege al controlador SDN 29 de la sobrecarga del plano de datos, lo que le permite prestar servicio a otras aplicaciones SDN.

3. Detector de paquetes de flujo grande

El detector de paquetes de flujo grande 24 también es responsable de realizar un seguimiento de los nuevos flujos, que incluye la información de 5 tuplas, la duración y el volumen, mediante el uso de estructuras de datos eficientes en memoria. Si un flujo está activo por más de un volumen umbral, se considera un flujo de elefante, y el detector de flujo grande 24 informa al corredor 25, que luego realiza una llamada a la API RESTful al controlador SDN 29 para insertar una entrada de tabla de flujo reactivo correspondiente en el conmutador SDN 23. Esto suprime el tráfico del plano de datos para este flujo de ser replicado en el detector de flujo grande 24, y también activa la telemetría para ese flujo de elefante.

La otra responsabilidad del detector de flujo grande 24 es la detección de las respuestas de tipo A de DNS, sobre las que extrae el nombre de dominio y las direcciones IP del servidor, y envía estos a través de JSON al intermediario de datos 25, que lo escribe en una tabla de base de datos DNS de series temporales de la base de datos 27. Esta

base de datos 27 se usa para asociar cada flujo de video con su proveedor de contenido.

4. Procedimiento de telemetría

El intermediario de datos 25 consulta estadísticas por flujo (contadores), las almacena en una tabla de base de datos de flujo de series temporales ("Flujo DB" en la Figura 1(b)) con información de marca de tiempo que representa una marca de tiempo correspondiente de cada consulta (*por ejemplo.*, la hora actual), y expone los datos almacenados a la interfaz de usuario a través de las API RESTful apropiadas. La telemetría recoge estadísticas de uso por flujo (grano fino) y por grupo (grano grueso) mediante el uso del módulo recolector de estadísticas de nuestra aplicación SDN.

4.1 Identificación de Vídeo: De acuerdo con la discusión anterior, el detector de flujo grande 24 identifica todos los flujos de elefante, que pueden incluir una mezcla de flujos de vídeo y otros flujos de elefante, y luego evita que sus paquetes se reflejen.

Se ejecuta un procedimiento de identificación de video para distinguir las transmisiones de video de las transferencias de elefante e identificar sus proveedores de contenido y resoluciones. A un alto nivel, el procedimiento de identificación de vídeo: (a) determina los atributos de un flujo dado, que luego se alimentan en un clasificador inteligente para distinguir los flujos de vídeo de las transferencias de elefante, (b) consulta la base de datos DNS ("DB DNS" en la Figura 1(b)) mediante el uso de la dirección IP del cliente/servidor del flujo para asociar el flujo de vídeo con su proveedor de contenido, y (c) estima la resolución del flujo de vídeo (en la realización descrita, la resolución se estima como una de Baja, Media, Alta o Ultra Alta).

4.2 Recolección y almacenamiento del uso: El intermediario de datos 25 recopila los datos del contador que representan los contadores de flujo por proveedor de contenido (tabla de grupo) y la entrada de la tabla de flujo reactivo de la transmisión de vídeo por cada entrada de la tabla de flujo reactivo. Mientras que el número de entradas en la tabla de grupos es generalmente relativamente pequeño y fijo, el número de entradas de flujo reactivo puede variar significativamente con el tiempo. La consulta del último cuando el número de entradas es grande puede dar como resultado una respuesta de múltiples partes, por ejemplo, un conmutador SDN Noviflow 23 rompe la respuesta en fragmentos de 2.500 flujos cada uno, lo que pone una presión considerable sobre el agente en el conmutador 23 y, en consecuencia, afecta la oportunidad de los resultados. Para mitigar este efecto, en la realización descrita el aparato 20 ajusta la frecuencia de sondeo en función del número de entradas en la tabla de flujo reactivo. Específicamente, cuando el número de entradas de flujo reactivo es menor que 2.500, el aparato 20 consulta los contadores cada segundo, pero reduce la frecuencia de consulta a una vez cada 4 segundos cuando el número de entradas excede 10.000. Cuando el intermediario de datos 25 almacena los datos del contador recibidos desde el conmutador SDN 23, almacena los datos del contador recibidos junto con la información de la marca de tiempo correspondiente de manera que puedan generarse perfiles de flujo que representan las características temporales de cada flujo. Los contadores de flujo/nivel de grupo se almacenan por tanto en una base de datos de flujo de series temporales, como se muestra en la Figura 1(b), y se envían periódicamente en un mensaje con formato JSON a un procedimiento de aprendizaje automático del intermediario de datos 25 como se describe a continuación.

5. Clasificación mediante el uso del aprendizaje automático

El intermediario de datos 25 ejecuta un procedimiento de clasificación de aprendizaje automático para determinar si el tráfico que pertenece a un flujo es un video en directo o no (un procedimiento de "identificación de video") y, si es así, para determinar la resolución de la transmisión de video (un procedimiento de "clasificación de resolución").

5.1 Atributos: La selección de atributos es de suma importancia para entrenar los clasificadores, dado que los clasificadores deben ser predictivos para identificar y clasificar correctamente los flujos de video. La Figura 3 muestra gráficos de patrones de tráfico observados para varios flujos de video de diferentes proveedores de contenido, por ejemplo, Youtube™, Netflix™ y Twitch™ (en diferentes resoluciones: baja, media, alta y ultra alta definición), y otros flujos de elefantes que incluyen los de Facebook™ la aplicación y las descargas grandes (representativas de las transferencias masivas o GoogleDrive™ o Dropbox™ sincronización de almacenamiento en la nube) durante los primeros tres minutos de su actividad

Se puede ver que, debido al almacenamiento en memoria intermedia que acompaña a la transmisión de vídeo, la característica de tiempo inactivo (es decir, la fracción de tiempo en la que no se intercambia ningún dato) de los flujos de vídeo en las Figuras 3(a) a 3(f) es bastante distintiva en comparación con el flujo grande de descarga en la Figura 3(h)). La velocidad promedio (que se muestra mediante líneas rojas punteadas) del Youtube™ 2.160 p (video 4 k de ultra alta definición) en la Figura 3(d) es mucho mayor que el de otras resoluciones de video (que se muestran en las Figuras 3(a)-3(c) y 3(e)-3(f)), pero es comparable a la gran descarga en la Figura 3(h). Además del tiempo inactivo y la velocidad promedio, la característica de la explosividad de cada flujo también es distintiva: el video de baja resolución y la gran descarga exhiben los patrones más y menos explosivos respectivamente, entre estos perfiles representativos que se muestran en la Figura 3. En base a estas observaciones visuales, es evidente que el tiempo inactivo, la velocidad promedio y la explosividad son capaces de identificar y clasificar colectivamente los

flujos de video. Por ejemplo, el Facebook™ El flujo de la aplicación mostrado en la Figura 3(g) exhibe características similares de los flujos de video (mostrados en las Figuras 3(b)-3(c)) en términos de tiempo inactivo y ráfagas, pero su velocidad está muy por debajo de la de los flujos de video.

La velocidad promedio y la fracción de tiempo inactivo para un flujo se pueden calcular sobre una ventana móvil (de decir, un minuto). La explosividad del tráfico de flujo se puede calcular de varias maneras, y se observa (particularmente en la caracterización del tráfico dependiente de largo alcance) que debe medirse en múltiples escalas de tiempo. En consecuencia, en las realizaciones descritas se calcula un coeficiente de variación (es decir, la relación de la desviación estándar a la media, CV = σ/μ) para las transmisiones en las granularidades de tiempo de 1, 2, 4, 8 y 16 segundos para proporcionar los valores respectivos que se denominan en la presente memoria como σ_1/μ, σ_2/μ, σ_4/μ, σ_8/μ y σ_{16}/μ. Estas medidas de la velocidad de variación, además del tiempo inactivo y la velocidad promedio μ de cada flujo, se proporcionan como atributos a los clasificadores. Tenga en cuenta que, para un nuevo flujo, puede haber solo un subconjunto de atributos de la explosividad al principio, porque el cálculo de σ_{16} requeriría la recopilación de datos durante al menos un minuto. Un flujo que comenzó hace solo 20 segundos solo podría producir σ_1/μ, σ_2/μ y σ_4/μ ya que hay menos de 4 puntos de datos en escalas de tiempo de 8 segundos y 16 segundos.

5.2 Identificación/Clasificación

Como se describió anteriormente, el intermediario de datos 25 ejecuta dos clasificadores, específicamente, el identificador de video (para indicar si un flujo es un video en tiempo real o no) y el clasificador de resolución (para determinar la resolución de un flujo de video durante la reproducción). Cada clasificador se invoca periódicamente (cada 16 segundos en la realización descrita) - la invocación inicial puede tener acceso a solo cinco atributos (tiempo inactivo, μ, σ_1/μ, σ_2/μ y σ_4/μ), y las invocaciones posteriores que tienen acceso a más atributos (relacionados con la explosividad) pueden cambiar la clasificación, mejorando la precisión y/o identificando los cambios de resolución. El entrenamiento de los clasificadores se describe a continuación.

Ejemplo

Se construyó una realización del aparato mediante el uso de componentes de software de código abierto que se muestran en forma de diagrama de bloques en la Figura 4. Este aparato 40 identifica y clasifica los flujos de vídeo en tiempo real a velocidades de línea de hasta 10 Gbps. En esta realización, la aplicación de SDN se implementa sobre el controlador SDN de código abierto Ryu (como se describe en https://osrg.github.io/ryu/), aumentado por el motor de inspección de paquetes Bro de código abierto (https://www.bro.org/) para la gestión del estado del flujo y el desencadenamiento de eventos, y las bases de datos se generan mediante el uso de la plataforma de base de datos de series temporales InfluxDB (https://www.influxdata.com/), la base de datos relacional de código abierto PostgreSQL (https://www.postgresql.org/) y CouchDB (http://couchdb.apache.org/), y una interfaz gráfica de usuario web escrita mediante el uso de la biblioteca de interfaz gráfica de usuario Javascript ReactJS (https://reactjs.org/) para la interacción con el usuario. Además, cada uno de estos componentes se ejecuta en un contenedor de docker o máquina virtual (VM) separado en un entorno de nube proporcionado por el hipervisor VMware Esxi 6.0. Cada una de las VM ejecuta el sistema operativo Ubuntu server 14.04 LTS y se le asigna una CPU de cuatro núcleos con 8 GB de memoria y 32 GB de espacio en disco.

Este aparato 40 está gestionando actualmente tres entornos: (a) una red universitaria experimental habilitada para SDN que abarca varios puntos de acceso WiFi, (b) un enlace punto a punto sobre el cual un generador de tráfico Spirent a escala industrial alimenta el tráfico en nuestra configuración, y (c) un enlace de red de dormitorio del campus en vivo que opera a 10 Gbps y que sirve a varios cientos de usuarios reales.

6.1.1 Conmutador SDN: El conmutador SDN 41 es un NoviSwitch 2116 totalmente compatible con Openflow 1.3, como se muestra en la Figura 4. Proporciona 160 Gbps de rendimiento, decenas de miles de entradas de flujo TCAM y millones de entradas de flujo de coincidencia exacta en DRAM.

6.1.2 Detector de flujo grande: La herramienta de código abierto Bro (v2.4.1) 42 se usa para la inspección del tráfico replicado. Los controladores de eventos se escribieron en Bro para realizar un seguimiento de la duración y el volumen del flujo, y para activar una llamada a la API al corredor de datos cuando se detecta un flujo de elefante. De manera similar, las respuestas DNS también se analizan y la información se pasa al intermediario de datos 41 para su registro en la base de datos de series temporales.

6.1.3 Corredor de datos: El intermediario de datos 43 en esta realización se escribe en el lenguaje Python. El intermediario de datos 43 recibe el 5-tuple de flujos de elefante e información de DNS desde el detector de flujo grande Bro 42, inserta/modifica entradas de flujo/grupo y recopila datos estadísticos de la aplicación SDN 44 a través de una API RESTful. Las estadísticas de flujo y grupo recopiladas de la aplicación SDN 44 se escriben en una base de datos InfluxDB de series temporales 46. La información del nivel de flujo se consulta de la base de datos InfluxDB 46 periódicamente para su procesamiento por el clasificador inteligente impulsado por la herramienta Weka (v3.8) (como se describe en https://en.wikipedia.org/wiki/Weka_(machine_learning)) mediante el uso de la interfaz de envoltura de Python de Weka (v0.3.9). El clasificador inteligente identifica los flujos de video, consulta la base de

datos DNS para etiquetar los flujos de video, llama a las API RESTful para modificar el grupo de salida de las entradas de flujo e identifica las resoluciones de flujo de video.

6.1.4 Controlador y aplicación de SDN: En esta realización se usa un controlador Openflow Ryu (v4.0) 46. La aplicación SDN 44 se escribe en Python y expone las API RESTful de norte a sur al corredor de datos 43 para insertar o modificar las reglas de la red y las estadísticas del flujo de sondeo. Las llamadas de API RESTful exitosas dan como resultado acciones apropiadas (por ejemplo, inserción, modificación y recopilación de reglas de red) en el conmutador SDN 41 que sirve al plano de datos.

6.1.5 Bases de Datos: Hay tres bases de datos en el sistema 40 para almacenar estadísticas de uso del flujo, información DNS y configuraciones del sistema. La serie temporal InfluxDB (v1.0.0) 46 se usa para almacenar estadísticas de flujo/grupo periódicas. En el mismo InfluxDB 46, también se almacena la información de las respuestas de tipo A de DNS, que incluye el nombre de dominio y las direcciones IP del cliente/servidor. Se usa una base de datos relacional de objetos PostgreSQL (v9.6.3) para almacenar el mapeo entre las direcciones IP de dominio y el sufijo del nombre de dominio. Se usa una base de datos orientada a documentos CouchDB NoSQL (v2.0.0) para almacenar las configuraciones del conmutador SDN 41 tales como el ID de ruta de datos (DPID) de flujo abierto y las configuraciones de múltiples tablas.

6.1.6 Interfaz Web: El aparato 40 proporciona una interfaz gráfica de usuario interactiva (GUI) o 'interfaz frontal' 50 para que los administradores de red visualicen flujos de video en su red, implementado en ReactJS mediante el uso de la plantilla Rubix y la biblioteca D3. Las capturas de pantalla de ejemplo se muestran en las Figuras 5 y 6.

6.2 Entrenamiento de máquinas

Los clasificadores del aparato 40 se entrenaron con conjuntos de datos recopilados por el propio aparato 40. Con el fin de tener la verdad objetiva para el entrenamiento, se escribió una secuencia de comandos de Python para generar la transmisión de vídeo desde varios proveedores, específicamente, Youtube™, Netflix™, Youku™, Facebook™, Tencent™, y otro tráfico de larga duración, que incluye descargas grandes (por ejemplo, sincronización de Google-Drive) y páginas web dinámicas (es decir, Office 365, página de inicio de Facebook, WhatsApp), a través de una red SDN WiFi experimental llamada "uniwide_sdn". Se usó la API del reproductor de Youtube para transmitir videos en resoluciones especificadas, específicamente, baja: 144 p, 240 p, 360 p; media: 480 p, 720 p; alta: 1.080 p, 1.440 p; y ultra alta: 4 K.

Con el propósito de entrenar, las secuencias de comandos limitan cada flujo (video y no video) a 128 segundos (es decir, aproximadamente dos minutos), aunque cada video elegido tenía una longitud total superior a 20 minutos. Navegador de Internet Firefox™ se utilizó la versión 47.0 para reproducir los vídeos. Las secuencias de comandos reproducen videos de los 5 proveedores más populares, en diferentes resoluciones de video, así como también diferentes archivos ISO grandes para descargar y sincronizar con Google Drive, para diversificar los conjuntos de datos de entrenamiento.

Al final de cada actividad de dos minutos, la secuencia de comandos consultó la InfluxDB 46 para extraer el perfil de flujo (recuento de bytes en el intervalo de tiempo de 1 segundo) y calcular los atributos como se describió anteriormente. El perfil de tráfico de 128 segundos se dividió en 8 subperfiles (correspondientes a intervalos de tiempo de [1, 16] s, [1, 32] s, [1, 48] s, [1, 64] s, [17, 80] s, [33, 96] s, [48, 112] s y [65, 128] s). La secuencia de comandos finalmente calculó los atributos para cada uno de los subperfiles. Tenga en cuenta que los subperfiles cortos (por ejemplo, [1, 16]) tendrán atributos incompletos tales como σ_8/μ y σ_{16}/μ. La secuencia de comandos se ejecutó durante 2 semanas, que recolectó un total de 28.543 instancias de entrenamiento etiquetadas para flujos de elefantes (video y no video), de las cuales 10.416 instancias se etiquetaron por resolución de video.

La Figura 7 muestra los histogramas resultantes de cada atributo para el identificador de video, y las diferencias son visualmente evidentes. Por ejemplo, el histograma de tiempo inactivo en la Figura 7(a) muestra que los tiempos inactivos de los flujos que no son de vídeo se centran en aproximadamente 1 % con desviaciones menores, mientras que los tiempos inactivos de los flujos de tráfico de vídeo se distribuyen ampliamente entre el 20 % y el 95 %. Las transmisiones de vídeo y no vídeo no son muy distintas en su histograma de velocidad promedio en la Figura 7(b). Sin embargo, son bastante diferentes en su comportamiento de la explosividad en varias escalas de tiempo, como se observa en las Figuras 7(c)-7(g).

La Figura 8 muestra las distribuciones de atributos para el clasificador de resolución. Como se esperaba, a medida que la resolución aumenta de baja a ultra alta, la distribución de la velocidad promedio se desplaza a la derecha (Figura 8(b)), mientras que la distribución de la fracción de tiempo inactivo se desplaza a la izquierda (Figura 8(a)). La explosividad en diversas escalas de tiempo también disminuye, como se muestra en la Figura 8(g).

6.2.1 Validación Cruzada: La herramienta Weka se usó para entrenar y validar el procedimiento de aprendizaje automático para la identificación y clasificación de videos. Se emplearon tres algoritmos de clasificación populares, específicamente, J48, Bosque Aleatorio y MLP, que usan los atributos descritos anteriormente. La eficacia de los clasificadores se validó mediante el uso del procedimiento de validación cruzada de 10 veces.

El procedimiento de validación cruzada divide aleatoriamente el conjunto de datos en conjuntos de entrenamiento (90 % de las instancias totales) y de validación (10 % de las instancias totales). Esta validación cruzada se repite 10 veces. Los resultados se promedian para producir una sola métrica de rendimiento. La precisión del identificador de vídeo se muestra en forma de una matriz de confusión en la Figura 9. Más del 96 % de los flujos de vídeo se identifican correctamente mediante el uso de los algoritmos J48 y MLP, mientras que el bosque aleatorio tiene un rendimiento ligeramente peor. La identificación correcta de los flujos que no son de vídeo es superior al 92 % con J48, aunque el bosque aleatorio y el MLP funcionan peor. En general, el J48 da un rendimiento razonable, con falsos positivos (no video clasificado como video) por debajo del 8 % y falsos negativos (video clasificado como no video) por debajo del 4 %.

Las matrices de confusión para el clasificador de resolución se muestran en la Figura 10. Tanto J48 como bosque aleatorio producen una precisión general consistente de más del 98 %. Se observa que los vídeos de alta definición se clasifican erróneamente con más frecuencia que otras resoluciones y es más probable que se clasifiquen erróneamente como de resolución media. Sin sorpresa, los vídeos de baja resolución mal clasificados también tienen más probabilidades de etiquetarse como de resolución media. La geometría de las instancias de entrenamiento es más adecuada para los clasificadores basados en árboles de decisión (es decir, J48 y bosque aleatorio) que para los clasificadores basados en redes neuronales (es decir, MLP), lo que da como resultado una mejor precisión. Además, todos los atributos elegidos tienen contribuciones significativas para identificar/clasificar el tráfico de vídeo y dado que J48 usa un árbol de decisión para todas las instancias de entrenamiento, supera al bosque aleatorio que emplea una colección de árboles de decisión independientes, cada uno de los cuales considera un subconjunto aleatorio de instancias de entrenamiento.

Se usó Weka para evaluar el mérito promedio de cada atributo en el procedimiento de clasificación. La Figura 11(a) muestra que el tiempo inactivo y la explosividad a 2 segundos y 4 segundos ($\sigma 2/\mu$ y $\sigma 4/\mu$) son los atributos más importantes para identificar un flujo de video (mostrado por barras de golpe). Sin embargo, la velocidad promedio (μ) y el tiempo inactivo contribuyen más al clasificador de resolución.

La precisión del aprendizaje automático se evaluó mediante el uso de una combinación de instancias de varios subperfiles (desde los primeros 16 segundos hasta más de un minuto durante una vida útil de dos minutos). El rendimiento de los clasificadores para cada subperfil se estudió por separado. La Figura 11(b) sugiere que las secuencias de vídeo se identifican con una precisión de aproximadamente 60 % si solo los primeros 16 segundos de su perfil están disponibles para el clasificador. Se observa que el crecimiento en la longitud de los subperfiles mejora la precisión de manera significativa; después de 48 segundos, se logra una precisión del 90 %. De manera similar, la precisión del clasificador de resolución está altamente correlacionada con la longitud del subperfil, como se muestra en la Figura 11(c). Esto no es sorprendente, ya que varios atributos calculados durante los primeros 16 segundos no identifican/clasifican perfectamente los flujos de video debido a su almacenamiento en búfer inicial. Por ejemplo, un video de ultra alta resolución (Figura 3(d)) es muy similar a una gran descarga si el tiempo inactivo, la velocidad promedio y la explosividad se consideran solo para los primeros 16 o 32 segundos del perfil. Los atributos $\sigma 8/\mu$ y $\sigma 16/\mu$ se vuelven disponibles respectivamente solo después de 32 y 64 segundos de actividad de transmisión, y son bastante importantes para la clasificación.

6.2.2 Resumen: La identificación de flujos de video y sus resoluciones para flujos de elefante en base a sus características de nivel de flujo (en lugar de nivel de paquete) tales como tiempo inactivo, velocidad promedio y ráfagas en múltiples escalas de tiempo es factible en tiempo real. La Figura 11 confirma que el aparato 40 puede identificar correctamente los flujos de vídeo con aproximadamente 70 % de precisión dentro de los primeros 30 segundos, que aumenta a más del 95 % de precisión en dos minutos. De manera similar, la clasificación de resolución logra más del 80 % de precisión en 30 segundos, que aumenta a más del 99 % en dos minutos.

7 Resultados de la evaluación

7.1 Prueba de escalabilidad

En esta sección, la eficacia del sistema se divulga al estresarlo con un gran número de flujos emulados mediante el uso de un Telescopio muestra (por línea morada) una carga promedio de alrededor de 274,90 Mbps dentro de un segundo, que está muy cerca de la velocidad de 279,56 Mbps reportada por las estadísticas de Spirent (es decir, un error de menos de 1,7 %). Se observa que la velocidad de transferencia del tráfico replicado (mostrado por la línea amarilla) alcanza un máximo de 273,45 Mbps y cae a cero gradualmente en 210 segundos.

Esto no es sorprendente, porque el enfoque adoptado en el presente sistema solo necesita que se envíen al analizador de tráfico los primeros segundos de tráfico de cada nuevo flujo de vídeo para su inspección; a partir de ahí, se inserta una entrada de flujo reactivo para detener el espejeo de paquetes. La carga replicado se ve afectada directamente por la velocidad de llegada de nuevos flujos de video. Tras la inserción del flujo reactivo, no se réplica ningún paquete de ese flujo y nuestra aplicación consulta después los recuentos de bytes para monitorear la actividad del flujo.

Las estadísticas de Spirent revelaron que se transfirieron 10,48 GB de datos, corroborando de cerca con los 10,44

GB medidos por nuestra aplicación del sistema. De esto, 4,35 GB se reflejó en el detector de flujo grande 24, correspondiente a aproximadamente el 42 % del tráfico total. La Figura 14 muestra la detección de flujos de elefante por nuestro sistema, y las correspondientes entradas de flujo reactivo se envían a una velocidad de 152 flujos por segundo, lo que da como resultado que casi cero paquetes se envíen al detector de flujo grande de software aproximadamente 4 minutos después del experimento. La prueba de estrés tenía como objetivo validar la escalabilidad de nuestro sistema a un gran número de flujos activos (31.920) y una alta velocidad de nuevos flujos (280/segundo), asegurando que tanto el detector de flujo grande de software 24 como el conmutador Openflow 23 puedan seguir el ritmo. Se encontró que los despliegues descritos a continuación tenían requisitos mucho más bajos en términos de números de flujo activo y llegadas de nuevo flujo, aunque las velocidades de datos absolutas eran más altas.

7.2 Clasificación del tráfico de los dormitorios del campus

El aparato 40 también se probó durante varios meses en la red cableada de la universidad que sirve a cientos de estudiantes.

La siguiente discusión proporciona información sobre los patrones de visualización de videos en el dormitorio, que se refiere al mes desde el 1 de mayo de 2017 hasta el 31 de mayo de 2017. La Figura 15(a) muestra un gráfico circular de la fracción de transmisiones de los proveedores de contenido de vídeo más populares - no es inesperado que los proveedores de contenido de vídeo gratuitos (Youtube y Facebook) sean los más dominantes, con un 44 % y un 17 %, respectivamente. Curiosamente, el número de transmisiones de video desde la plataforma de juegos Twitch (3 %) supera el número de transmisiones de Netflix (2 %). Se observa que el 8 % de los flujos de video se obtienen de los servidores multimedia de Akamai (es decir, akamai.net y akamaiedge.net). Por último, el sistema permitió la identificación de muchos otros proveedores de video en la nube tales como Tencent, Youku, Amazon, Fastly, Alibaba, Shifen - estos se agrupan como "Otros" en la Figura 15(a) que contribuyen colectivamente al 23 % de los flujos de video en el dormitorio.

La Figura 16 representa la función de distribución acumulativa complementaria (CCDF) de la duración y la velocidad promedio de las transmisiones de video de 4 proveedores de contenido populares, incluido Facebook™, Youtube™, Twitch™ y Netflix™, durante mayo de 2017. Como se muestra en la Figura 16(a), los videos de Twitch y Netflix se reproducen durante más tiempo (con una duración promedio de aproximadamente 10 minutos), seguidos de los videos de Youtube y Facebook con duraciones promedio de aproximadamente 3,5 y 1,5 minutos respectivamente en el dormitorio. Teniendo en cuenta la velocidad promedio en la Figura 16(b), los videos de Twitch y Netflix normalmente consumen más ancho de banda que los videos de Youtube y Facebook: Twitch y Netflix usan en promedio 6,6 Mbps, mientras que esta medida es de 2,8 y 1,5 Mbps para Youtube y Facebook, respectivamente.

En la Figura 15(b) se muestra el patrón de consumo de video día a día durante el mes. Observaciones interesantes que surgen de esto son que hay una fluctuación sustancial en la proporción relativa de proveedores de video de un día para otro, y parecería que los residentes de la residencia universitaria tendieron a ver más videos de juegos de Twitch los fines de semana que los días de la semana. La Figura 15(c) muestra la fracción de flujos de video a diferentes resoluciones de forma horaria (promediada durante el mes de mayo de 2017). Sorprendentemente, la mayoría de los videos se reproducen a una resolución media y solo una pequeña fracción de los videos está en ultra alta resolución, a pesar de que la red de campus universitario tiene un ancho de banda abundante y rara vez experimenta congestión. Esto se debe a que la mayoría de las películas gratuitas (o videoclips largos) solo están disponibles en resolución media o menos (es decir, 144 p, 240 p, 360 p, 480 p y 720 p) en Youtube y Facebook.

Sin embargo, el número de flujos de video por hora, junto con la distribución de su calidad, da visibilidad a la transmisión de video en la red de la Universidad que no fue factible antes, y es muy apreciado por el personal de TI de la universidad que puede obtener informes semanales y mensuales directamente del aparato 40.

Las realizaciones descritas de la presente invención combinan juiciosamente la inspección de paquetes de software a nivel de flujo con la telemetría a nivel de flujo de hardware, junto con el aprendizaje automático, para identificar y clasificar los flujos de vídeo en tiempo real y a bajo costo.

Las realizaciones y ejemplos anteriores se han descrito en el contexto de las aplicaciones para identificar y clasificar los datos de vídeo que fluyen a través de una red. Sin embargo, debe entenderse que la invención no se limita a monitorear datos de video y puede usarse para monitorear otros tipos de datos de red.

Muchas modificaciones serán evidentes para los expertos en la técnica sin apartarse del ámbito de la presente invención como se define en las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un procedimiento de monitoreo del tráfico de red ejecutado por un sistema de monitoreo del tráfico de red (20) de una red de comunicaciones, el procedimiento que incluye:

recibir paquetes de datos replicados de un conmutador de flujo de red definida por software, SDN, (23) de una red de comunicaciones;
procesar los campos de encabezado de los paquetes de datos recibidos para identificar subconjuntos de los paquetes de datos como pertenecientes a los respectivos flujos de red;
detectar flujos grandes de red al determinar, para cada uno de los flujos de red, una cantidad acumulativa correspondiente de datos contenidos en los paquetes recibidos para el flujo de red hasta que la cantidad acumulativa de datos alcance o supere una cantidad umbral predeterminada de datos;
para cada flujo grande de red detectado, enviar datos de identificación de flujo a la conmutación de flujo SDN (23) para permitir que la conmutación de flujo SDN (23) identifique los paquetes adicionales del flujo grande de red como paquetes del flujo grande de red y para detener la réplica de los paquetes adicionales del flujo grande de red a un componente del sistema de monitoreo del tráfico de red;
para cada flujo grande de red, recibir periódicamente desde el conmutador de flujo SDN (23) los datos del contador correspondientes que representan una cantidad acumulativa correspondiente de datos contenidos en los paquetes del flujo grande;
para cada flujo grande de red, procesar los datos del contador correspondientes y los datos de la marca de tiempo correspondiente para generar métricas temporales del flujo grande de red; y
para cada flujo grande de red, procesar las métricas temporales generadas con un clasificador entrenado para clasificar el flujo grande de red como uno de una pluralidad de tipos de flujo predeterminados, y en el que
las métricas de flujo incluyen tiempo inactivo, velocidad promedio y métricas de ráfagas.

2. El procedimiento de la reivindicación 1, en el que los tipos de flujo incluyen flujos de vídeo y flujos de no vídeo.

3. El procedimiento de la reivindicación 2, en el que los tipos de flujo incluyen flujos de video de respectivas resoluciones diferentes.

4. El procedimiento de cualquiera de las reivindicaciones 1 a 3, que incluye determinar los proveedores de servicios de al menos algunos de los flujos grandes de red de la información de DNS.

5. El procedimiento de cualquiera de las reivindicaciones 1 a 4, en el que las métricas de la ráfaga están en respectivas escalas de tiempo.

6. El procedimiento de la reivindicación 5, en el que las escalas de tiempo representan una serie geométrica.

7. Un procedimiento de monitoreo del tráfico de red ejecutado por un conmutador de flujo de red definida por software, SDN, (23) de una red de comunicaciones, el procedimiento que incluye las etapas de:

recibir un paquete de datos de la red de comunicaciones;
procesar el paquete de datos recibido para determinar si el paquete de datos es un paquete de una pluralidad de flujos grandes de red predeterminados, y si es así, identificar uno correspondiente de los flujos grandes de red predeterminados;
sí dicho procesamiento identifica un flujo grande de red correspondiente de paquete de datos predeterminado, entonces actualiza los datos del contador correspondiente que representan una cantidad acumulativa correspondiente de datos del flujo grande de red;
de lo contrario, si no se determina que el paquete de datos es un paquete de la pluralidad de flujos grandes de red predeterminados, entonces replicar el paquete de datos a un componente (24) de un sistema de monitoreo del tráfico de red (20) para determinar si el paquete de datos es un paquete de un flujo grande de red que no es uno de los flujos grandes de red predeterminados;
recibir datos de identificación de flujo grande de un componente (28, 29) del sistema de monitoreo del tráfico de red, identificando los datos de identificación de flujo grande al menos un flujo grande de red adicional que no es uno de los flujos de red predeterminados;
procesar los datos de identificación de flujo grande recibidos para añadir el al menos un flujo grande de red adicional a los flujos grandes de red predeterminados de manera que la etapa de procesamiento determinará que los paquetes de datos adicionales del al menos un flujo grande de red adicional son paquetes del al menos un flujo grande de red adicional y, en consecuencia, la conmutación de flujo de SDN (23) no replicará el paquete de datos al sistema de monitoreo del tráfico de red; y
enviar periódicamente, a un componente (25) del sistema de monitoreo del tráfico de red, datos del contador que representan las cantidades acumulativas correspondientes de datos contenidos en los flujos respectivos de los flujos grandes de red predeterminados.

8. Al menos un medio de almacenamiento legible por ordenador que tiene almacenadas en el mismo las

instrucciones ejecutables que, cuando se ejecutan por uno o más procesadores, hacen que uno o más procesadores ejecuten el procedimiento de cualquiera de las reivindicaciones 1 a 7.

9. Un sistema de monitoreo del tráfico de red (20), que incluye:

un componente de detección de flujo grande (24) que se configura para:

(i) recibir paquetes de datos replicados desde un conmutador de flujo de red definida por software, SDN, (23);
(ii) procesar los campos de encabezado de los paquetes de datos recibidos para identificar subconjuntos de los paquetes de datos como pertenecientes a los respectivos flujos de red;
(iii) detectar flujos grandes de red al determinar, para cada uno de los flujos de red, una cantidad acumulada correspondiente de datos contenidos en los paquetes recibidos para el flujo de red hasta que la cantidad acumulada de datos alcance o supere una cantidad umbral predeterminada de datos;
(iv) para cada flujo grande de red detectado, enviar datos de identificación de flujo al conmutador de flujo SDN (23) para permitir que el conmutador de flujo SDN (23) identifique los paquetes adicionales del flujo grande de red como paquetes del flujo grande de red y detener la réplica de los paquetes adicionales del flujo grande de red al componente de detección de flujo grande (24); y

un componente de análisis de flujo configurado para:

(i) para cada flujo grande de red, recibir periódicamente desde el conmutador de flujo de SDN (23) los datos del contador correspondientes que representan una cantidad acumulativa correspondiente de datos contenidos en los paquetes del flujo grande;
(ii) para cada flujo grande de red, procesar los datos del contador correspondientes y los datos de la marca de tiempo correspondiente para generar métricas temporales del flujo grande de red; y
(iii) para cada flujo grande de red, procesar las métricas temporales generadas con un clasificador entrenado para clasificar el flujo grande de red como uno de una pluralidad de tipos de flujo predeterminados,
y en el que las métricas de flujo incluyen tiempo inactivo, velocidad promedio y métricas de la ráfaga.

10. El sistema de la reivindicación 9, que incluye, además, un componente de interfaz de usuario (26) configurado para recibir las solicitudes de los usuarios y, en respuesta a las solicitudes, generar los datos de la interfaz de usuario que representan una interfaz de usuario interactiva para visualizar la información en flujos grandes de red detectados por el sistema, la información que incluye las clasificaciones de los flujos grandes de red.

11. El sistema de la reivindicación 9 o 10, que incluye, además, un conmutador de flujo de red definida por software, SDN (23) configurado para:

recibir un paquete de datos de la red de comunicaciones;
procesar el paquete de datos recibido para determinar si el paquete de datos es un paquete de una pluralidad de flujos grandes de red predeterminados, y si es así, identificar uno correspondiente de los flujos grandes de red predeterminados;
sí dicho procesamiento identifica un flujo grande de red correspondiente de paquete de datos predeterminado, entonces actualizar los datos del contador correspondiente que representan una cantidad correspondiente de datos del flujo grande de red;
de lo contrario, si no se determina que el paquete de datos es un paquete de la pluralidad de flujos grandes de red predeterminados, entonces replicar el paquete de datos a un componente (24) del sistema de monitoreo del tráfico de red para determinar si el paquete de datos es un paquete de un flujo grande de red que no es uno de los flujos grandes de red predeterminados;
recibir datos de identificación de flujo grande de un componente (28,29) del sistema de monitoreo del tráfico de red, identificando los datos de identificación de flujo grande al menos un flujo grande de red adicional que no es uno de los flujos de red predeterminados;
procesar los datos de identificación del flujo grande recibido para añadir el al menos un flujo grande de red adicional a los flujos grandes de red predeterminados de manera que la etapa de procesamiento determinará que los paquetes de datos adicionales del al menos un flujo grande de red adicional son paquetes del al menos un flujo grande de red adicional y, en consecuencia, la conmutación de flujo de SDN (23) no replicará el paquete de datos al componente del sistema de monitoreo del tráfico de red; y
enviar periódicamente, a un componente (25) del sistema de monitoreo del tráfico de red, datos del contador que representan cantidades acumulativas de datos contenidos en los flujos respectivos de los flujos grandes de red predeterminados.

12. El sistema de cualquiera de las reivindicaciones 9 a 11, en el que los tipos de flujo incluyen flujos de vídeo y flujos que no son de vídeo, y opcionalmente los tipos de flujo incluyen flujos de vídeo de respectivas resoluciones diferentes.

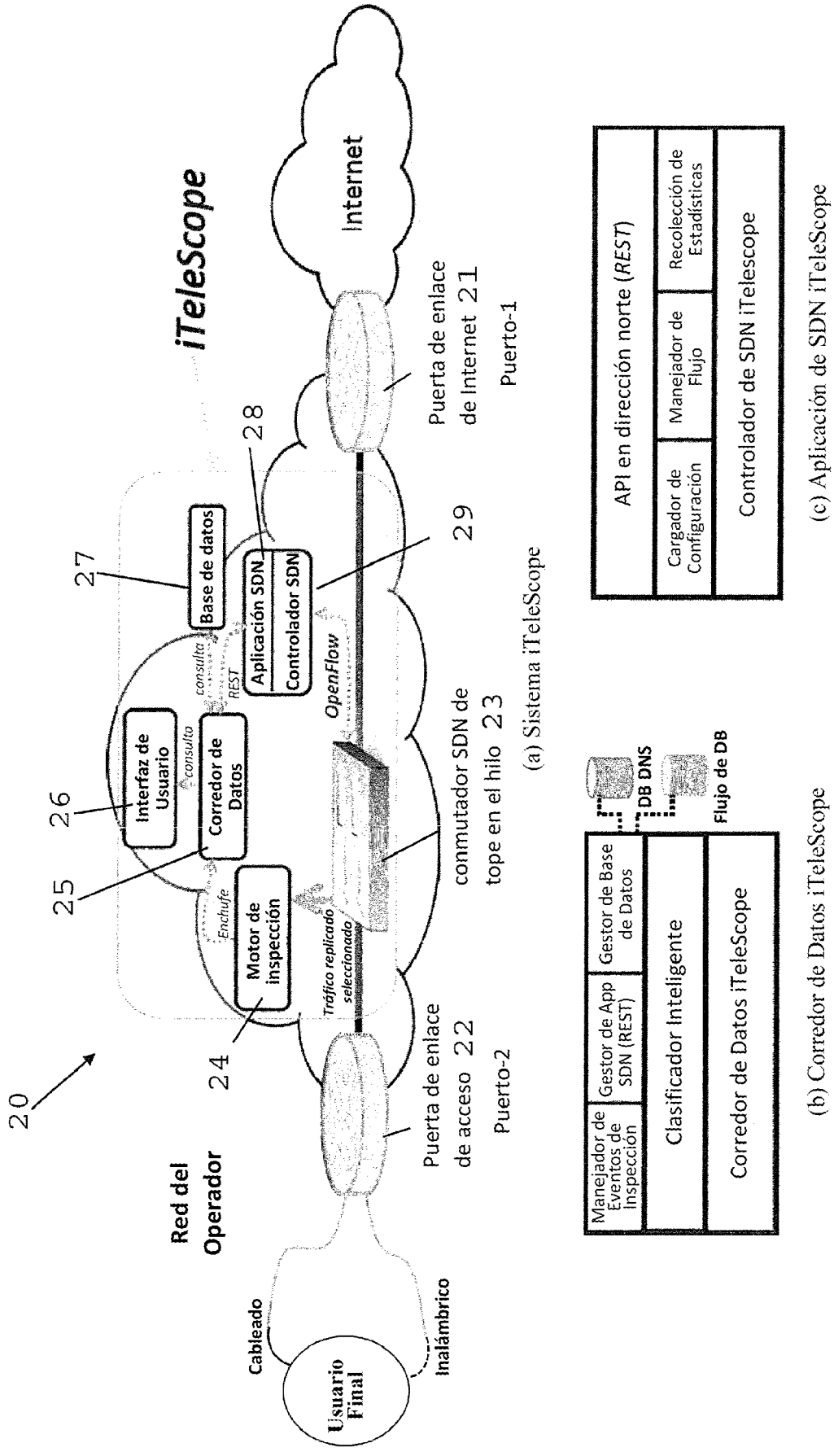

20
Red del Operador
iTeleScope
Internet
Usuario Final
Cableado
Inalámbrico
Puerta de enlace de acceso 22
Puerto-2
Puerta de enlace de Internet 21
Puerto-1
conmutador SDN de tope en el hilo 23
Interfaz de Usuario
Corredor de Datos
Motor de inspección
Base de datos
Aplicación SDN
Controlador SDN
consulta
REST
Enchufe
Tráfico replicado seleccionado
OpenFlow
24
25
26
27
28
29
(a) Sistema iTeleScope
Manejador de Eventos de Inspección
Gestor de App SDN (REST)
Gestor de Base de Datos
DB DNS
Clasificador Inteligente
Flujo de DB
Corredor de Datos iTeleScope
(b) Corredor de Datos iTeleScope
API en dirección norte (REST)
Cargador de Configuración
Manejador de Flujo
Recolección de Estadísticas
Controlador de SDN iTelescope
(c) Aplicación de SDN iTeleScope

FIGURA 1

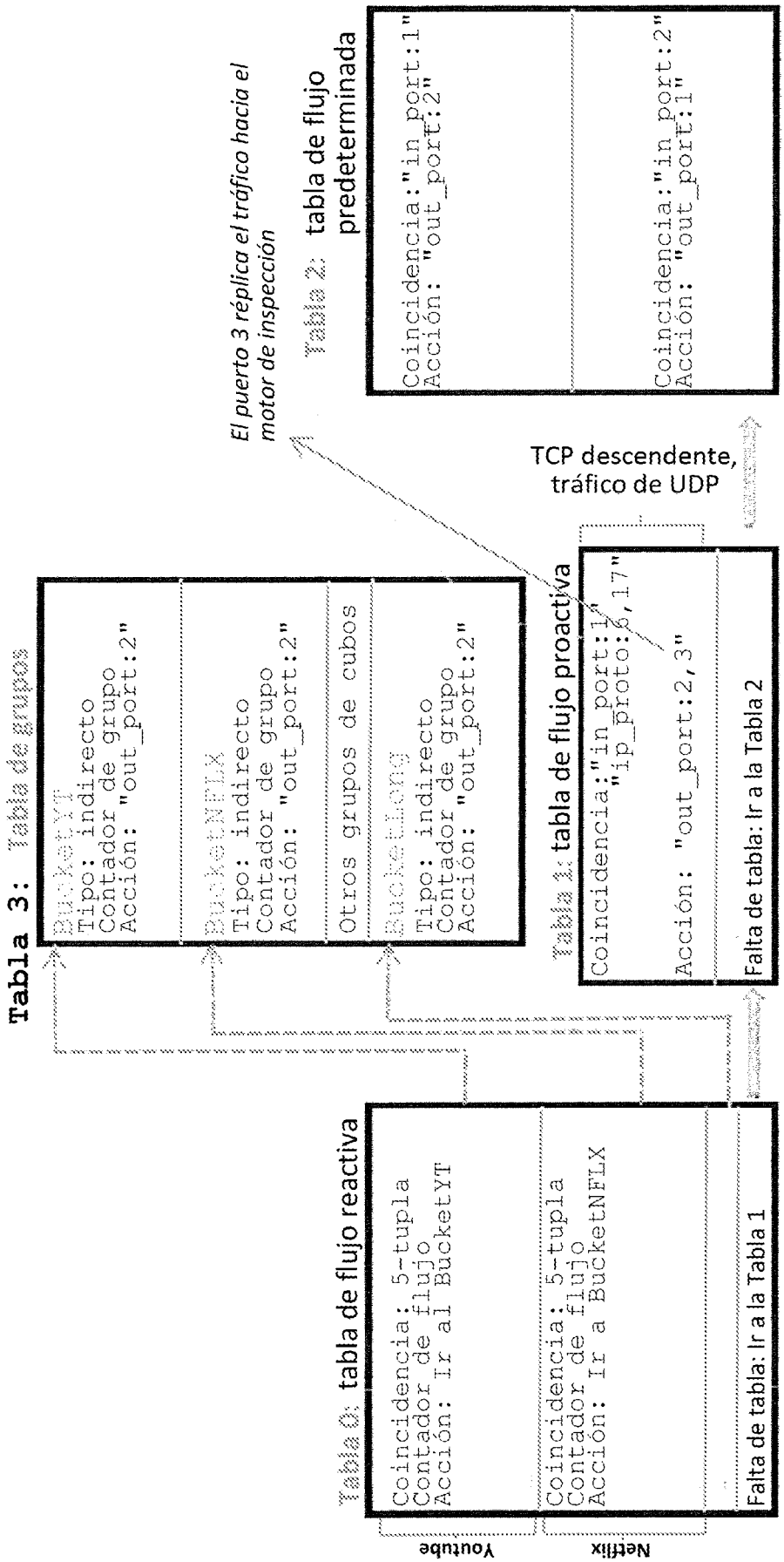
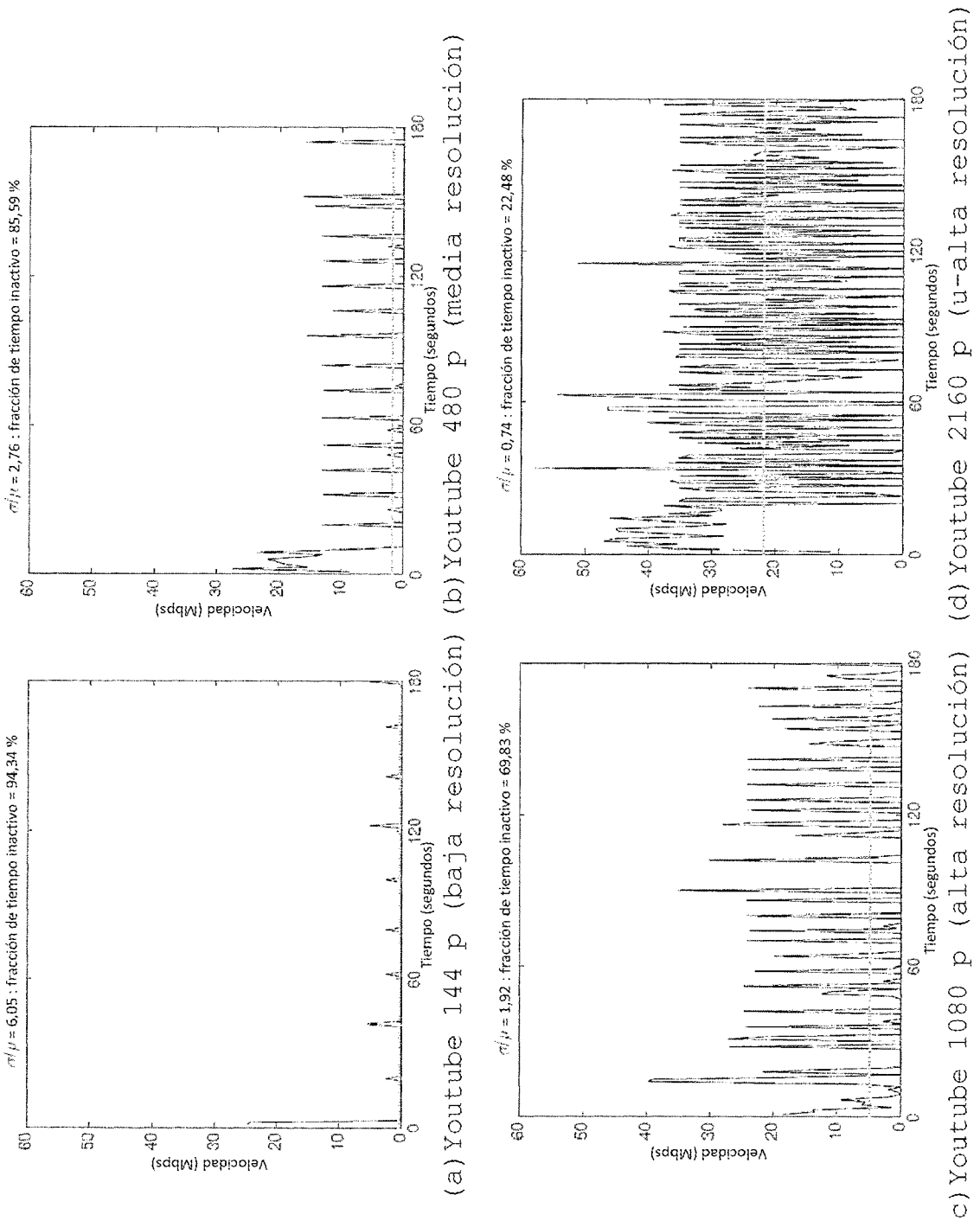


Tabla 0: tabla de flujo reactiva
Coincidencia: 5-tupla
Contador de flujo
Acción: Ir al BucketYT
Coincidencia: 5-tupla
Contador de flujo
Acción: Ir a BucketNFLX
Falta de tabla: Ir a la Tabla 1
Youtube
Netflix
Tabla 3: Tabla de grupos
BucketYT
Tipo: indirecto
Contador de grupo
Acción: "out_port:2"
BucketNFLX
Tipo: indirecto
Contador de grupo
Acción: "out_port:2"
Otros grupos de cubos
BucketLONG
Tipo: indirecto
Contador de grupo
Acción: "out_port:2"
Tabla 1: tabla de flujo proactiva
Coincidencia:"in_port:1"
"ip_proto:6,17"
Acción: "out_port:2,3"
Falta de tabla: Ir a la Tabla 2
TCP descendente,
tráfico de UDP
El puerto 3 réplica el tráfico hacia el
motor de inspección
Tabla 2: tabla de flujo
predeterminada
Coincidencia:"in_port:1"
Acción: "out_port:2"
Coincidencia:"in_port:2"
Acción: "out_port:1"

FIGURA 2



(a) Youtube 144 p (baja resolución) (b) Youtube 480 p (media resolución)

(c) Youtube 1080 p (alta resolución) (d) Youtube 2160 p (u-alta resolución)

FIGURA 3

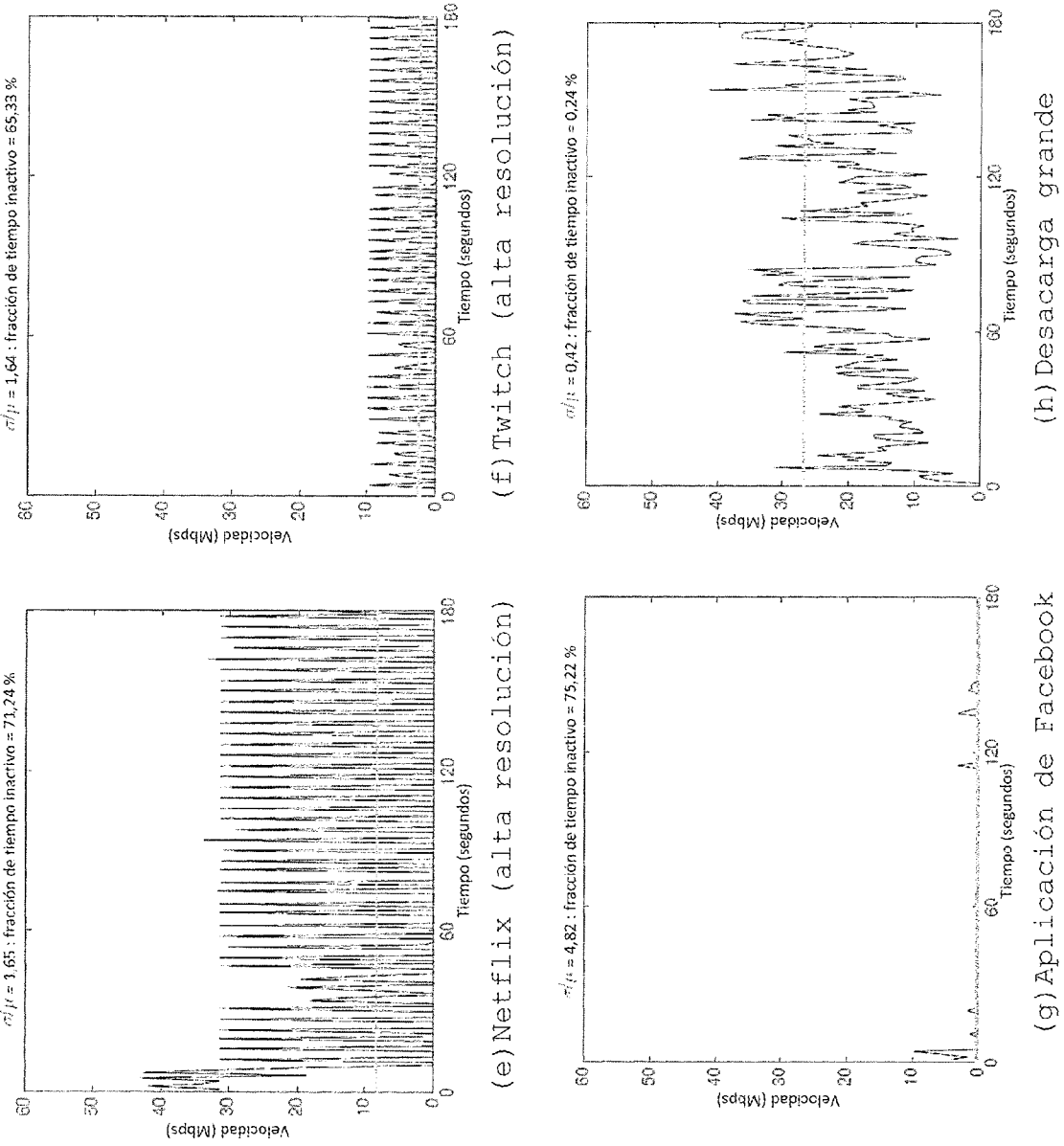

σ/μ = 1,65 : fracción de tiempo inactivo = 71,24 %
σ/μ = 1,64 : fracción de tiempo inactivo = 65,33 %
σ/μ = 4,82 : fracción de tiempo inactivo = 75,22 %
σ/μ = 0,42 : fracción de tiempo inactivo = 0,24 %
Velocidad (Mbps)
Tiempo (segundos)
0
10
20
30
40
50
60
120
180
(e) Netflix (alta resolución)
(f) Twitch (alta resolución)
(g) Aplicación de Facebook
(h) Desacarga grande

FIGURA 3

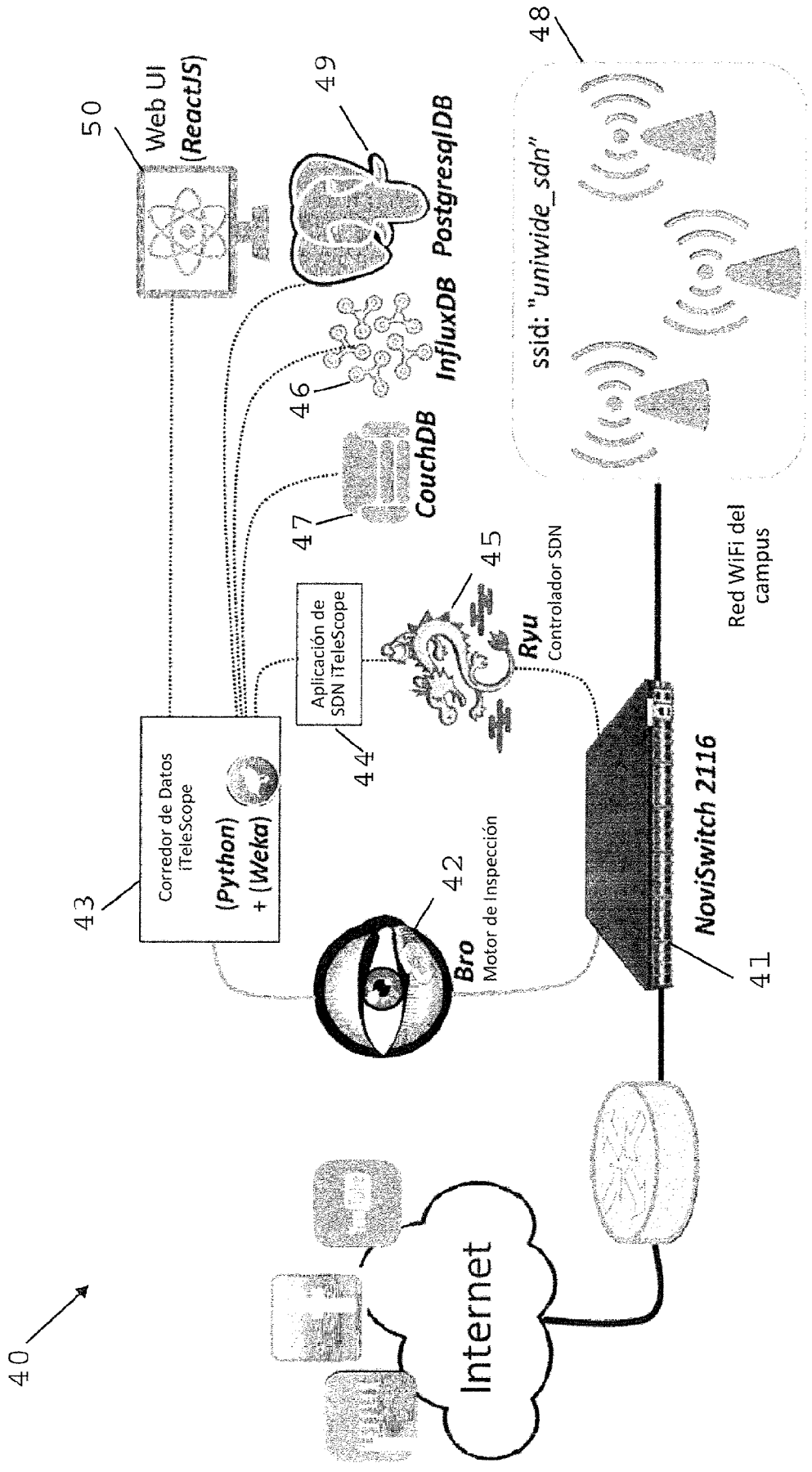

40
Internet
41
NoviSwitch 2116
Red WiFi del campus
48
ssid: "uniwide_sdn"
42
Bro
Motor de Inspección
43
Corredor de Datos iTeleScope
(Python)
+ (Weka)
44
Aplicación de SDN iTeleScope
45
Ryu
Controlador SDN
46
InfluxDB
47
CouchDB
49
PostgresqlDB
50
Web UI
(ReactJS)

FIGURA 4

Flujos Recientes

Etiqueta	IP del Servidor	IP del Cliente	Total de Bytes (MB)	Velocidad Estimada (Mbps)	Calidad	Duración (seg)
netflix	138.44.10.30	129.94.5.80	9,21	1,04	Medio	136
youtube	203.5.76.207	129.94.5.111	5,58	0,15	Bajo	228
netflix	138.44.10.30	129.94.5.84	219,58	7,69	Alto	259
youtube	203.5.76.206	129.94.5.80	910,10	17,00	Ultra-alto	431
facebook	157.240.8.23	129.94.5.80	131,62	3,12	Medio	535

FIGURA 5

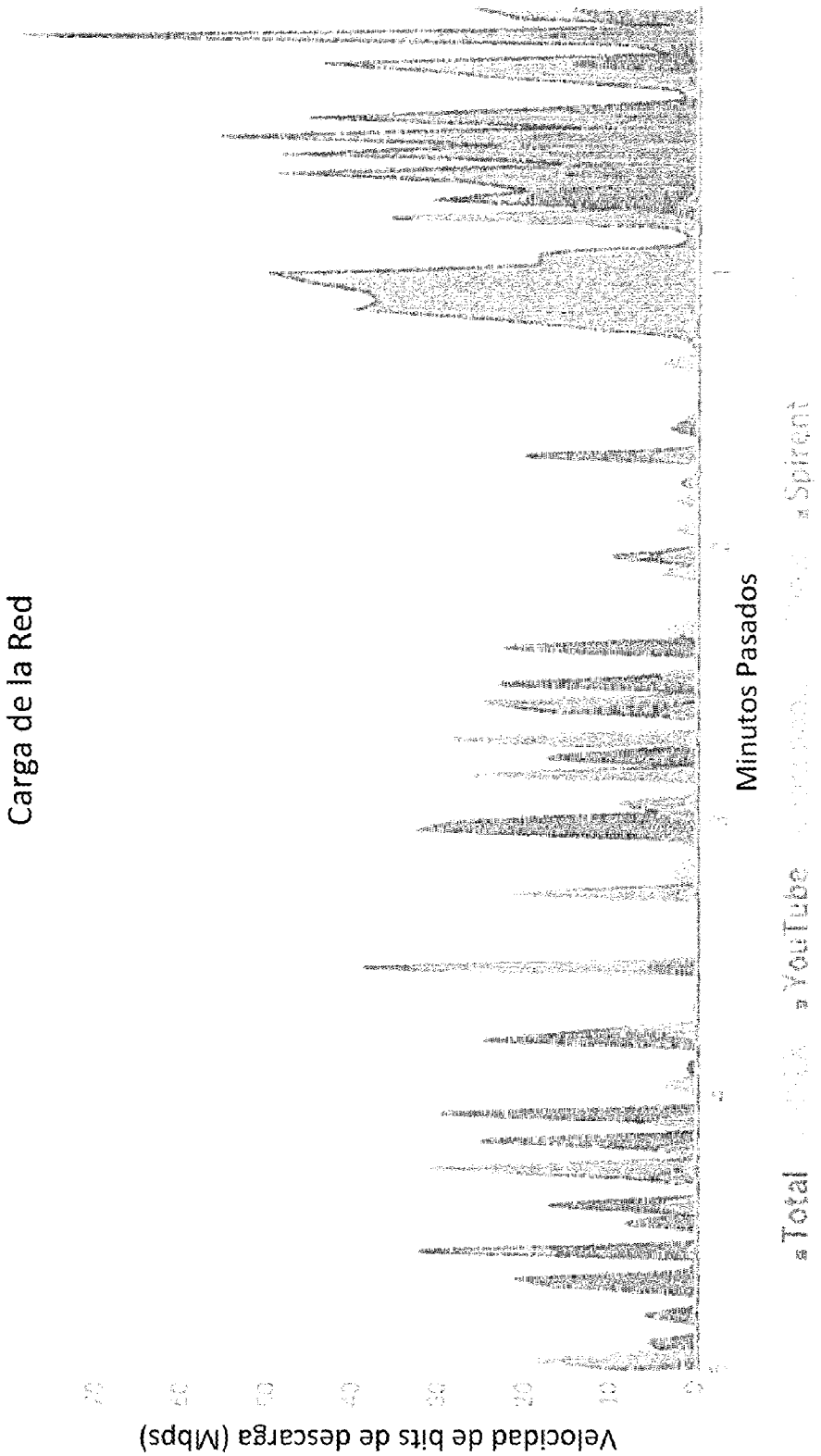
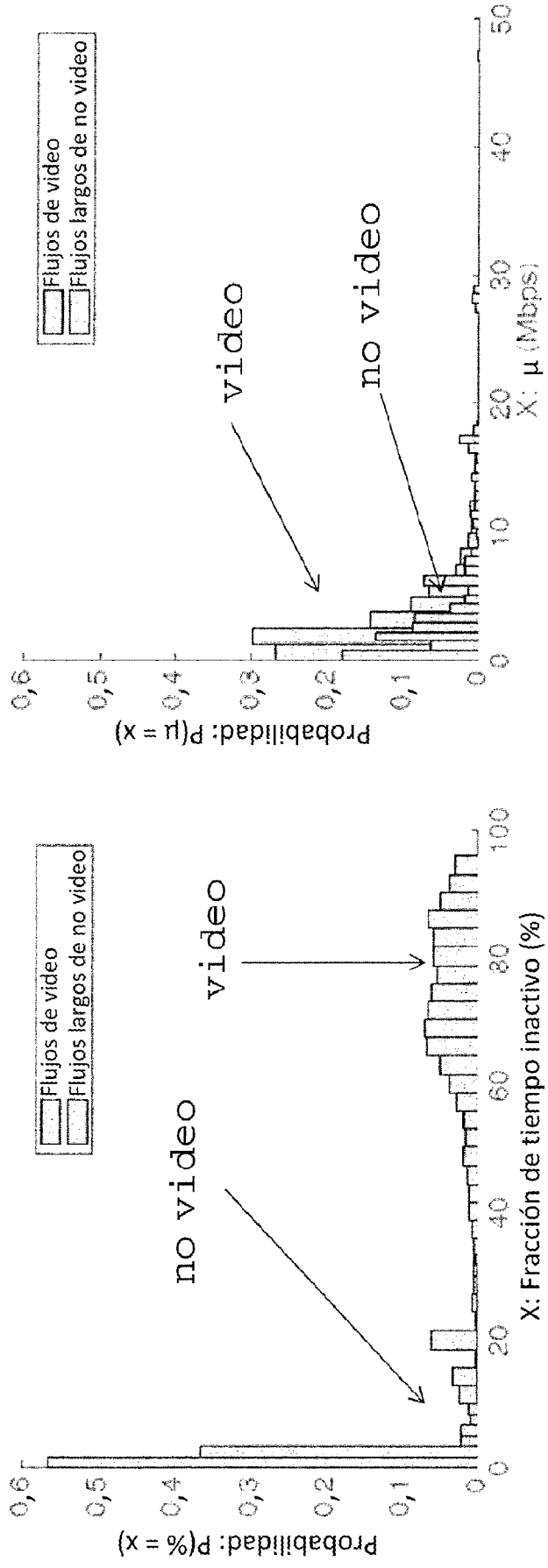

Carga de la Red
Minutos Pasados
Velocidad de bits de descarga (Mbps)

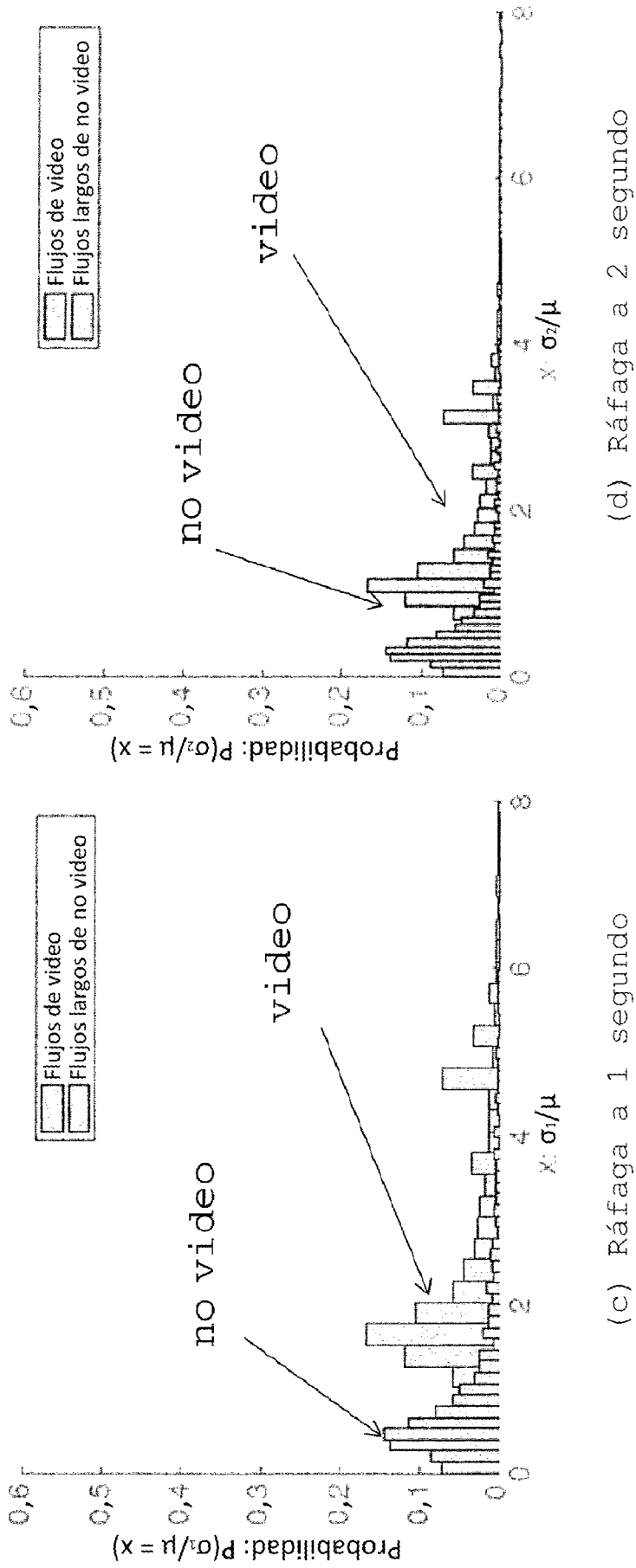
FIGURA 6



(a) Tiempo inactivo

(b) Velocidad promedio

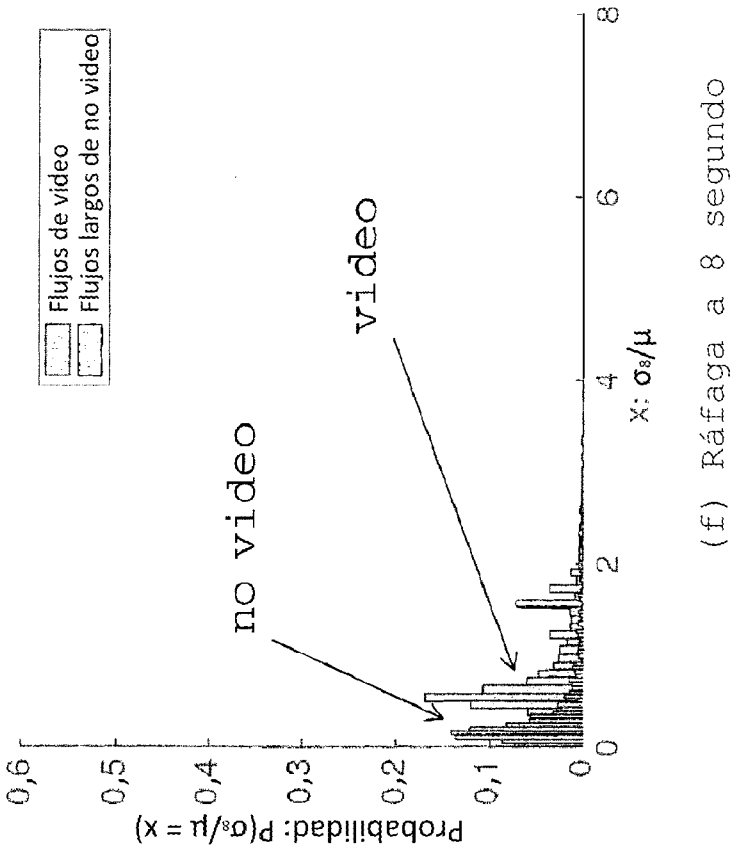
FIGURA 7



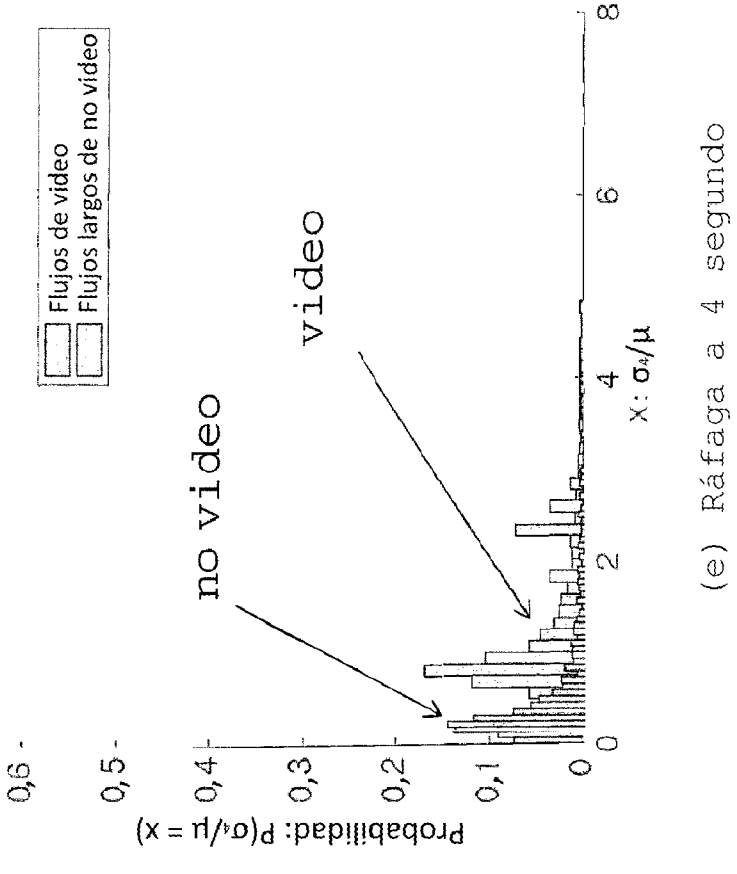
(c) Ráfaga a 1 segundo

(d) Ráfaga a 2 segundo

FIGURA 7

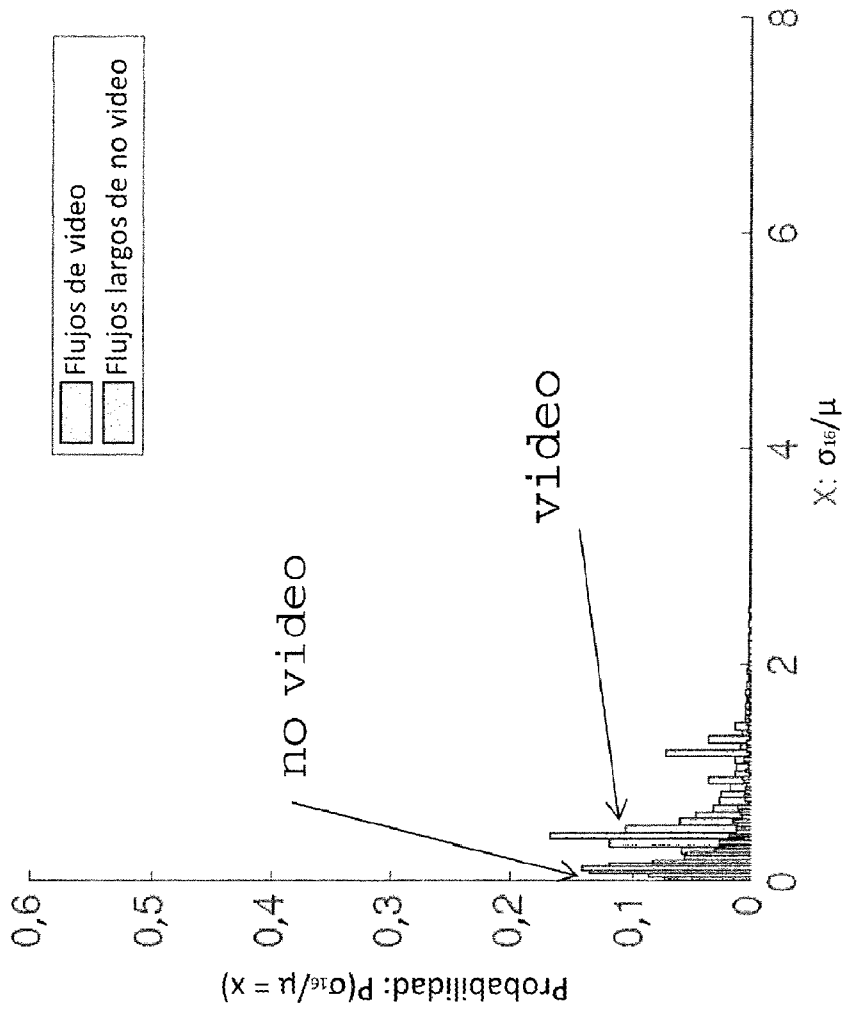


(f) Ráfaga a 8 segundo



(e) Ráfaga a 4 segundo

FIGURA 7



(g) Ráfaga a 16 segundo

FIGURA 7

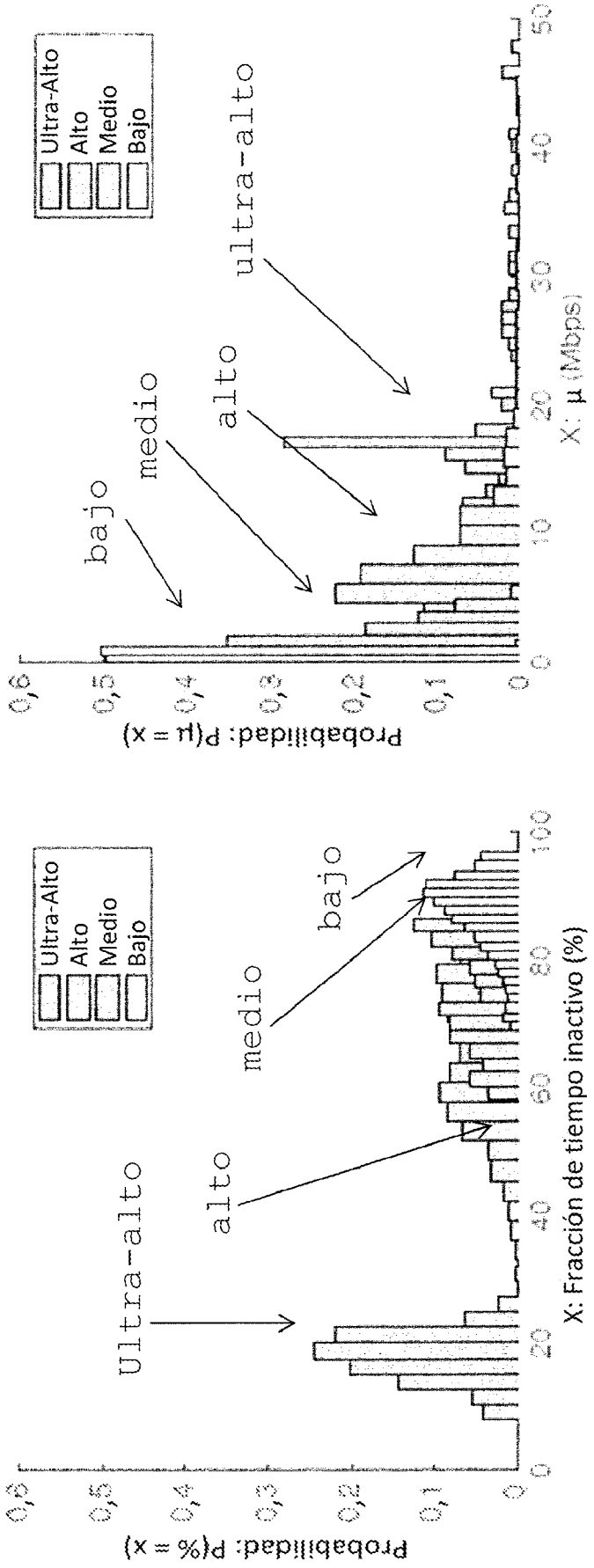

Ultra-Alto
Alto
Medio
Bajo
Ultra-alto
alto
medio
bajo
Probabilidad: P(% = x)
X: Fracción de tiempo inactivo (%)
0,6
0,5
0,4
0,3
0,2
0,1
0
20
40
60
80
100
(a) Tiempo inactivo
Ultra-Alto
Alto
Medio
Bajo
ultra-alto
alto
medio
bajo
Probabilidad: P(μ = x)
X: μ (Mbps)
0
10
20
30
40
50
(b) Velocidad promedio

FIGURA 8

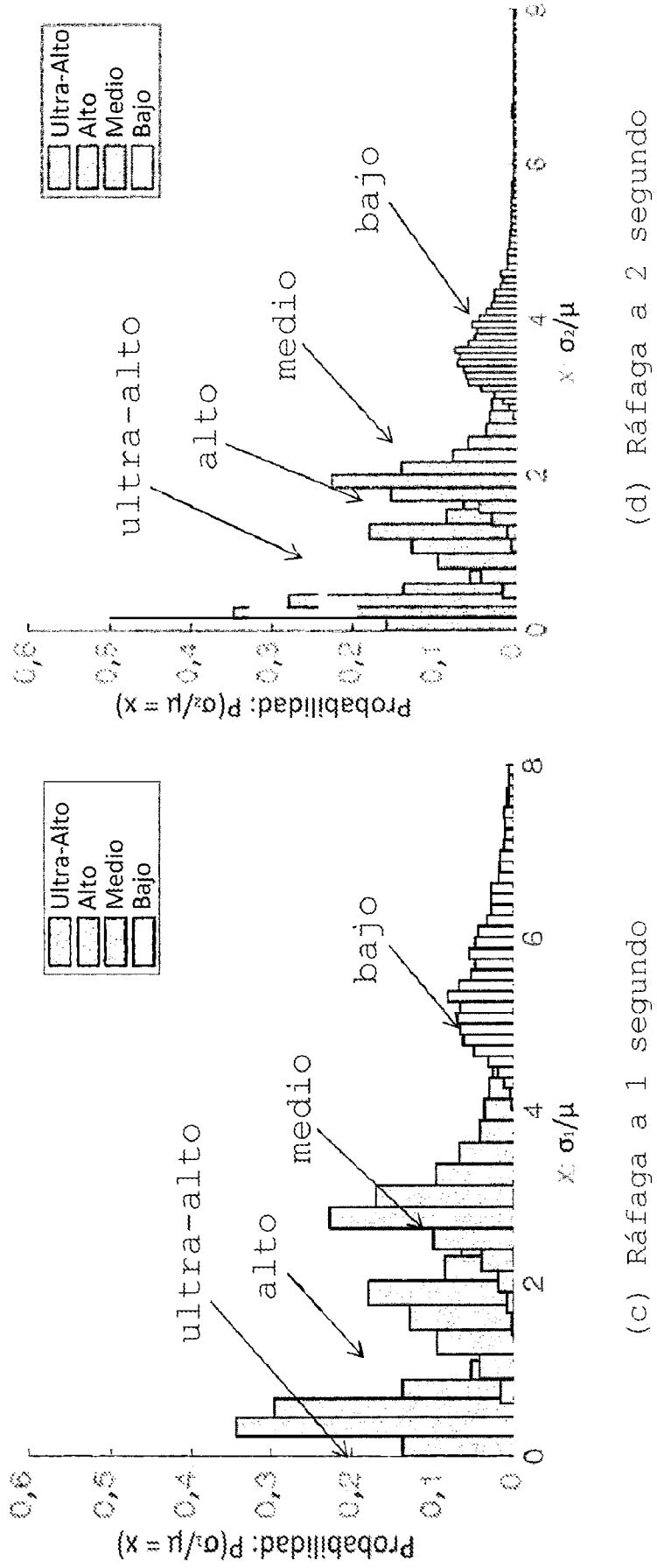

Ultra-Alto
Alto
Medio
Bajo
ultra-alto
alto
medio
bajo
Probabilidad: P(σ₂/μ = x)
x σ₂/μ
(d) Ráfaga a 2 segundo
Ultra-Alto
Alto
Medio
Bajo
ultra-alto
alto
medio
bajo
Probabilidad: P(σ₁/μ = x)
x σ₁/μ
(c) Ráfaga a 1 segundo

FIGURA 8

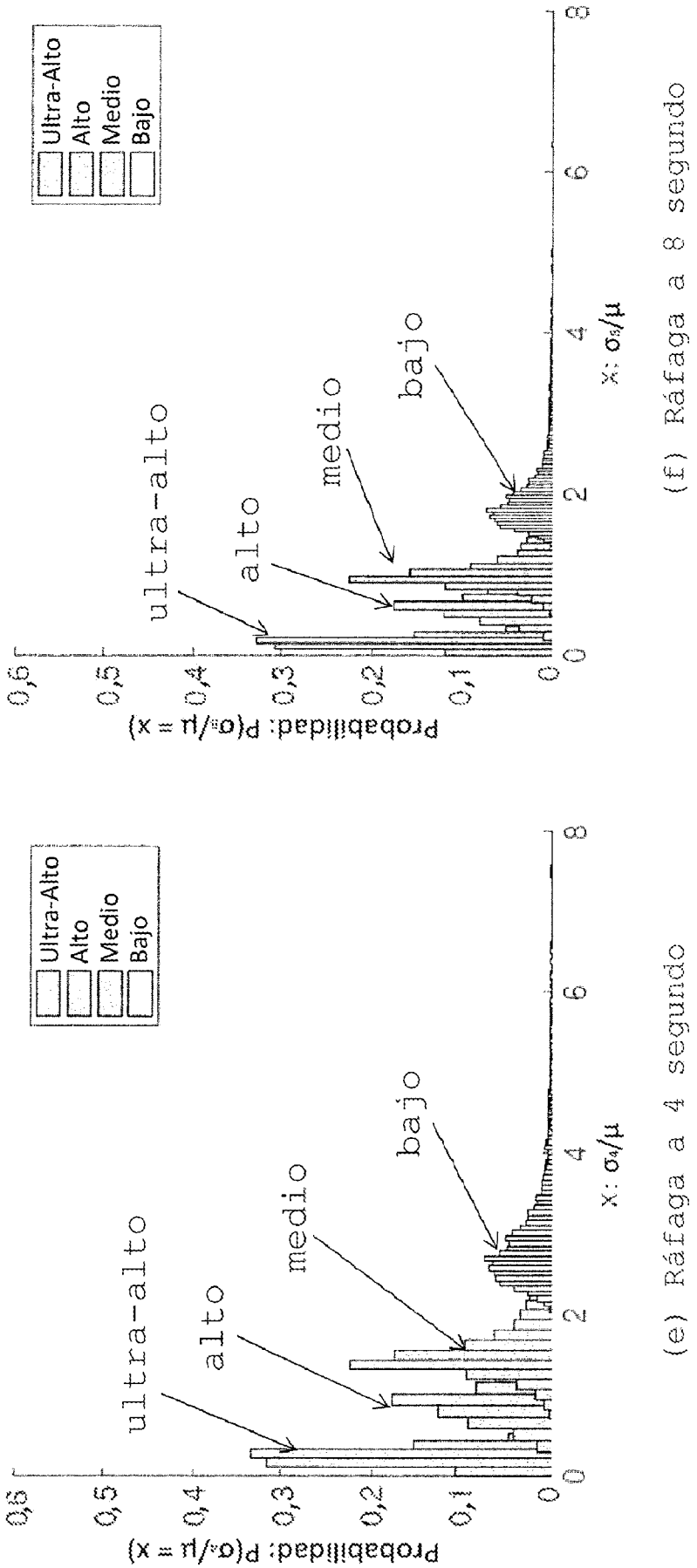
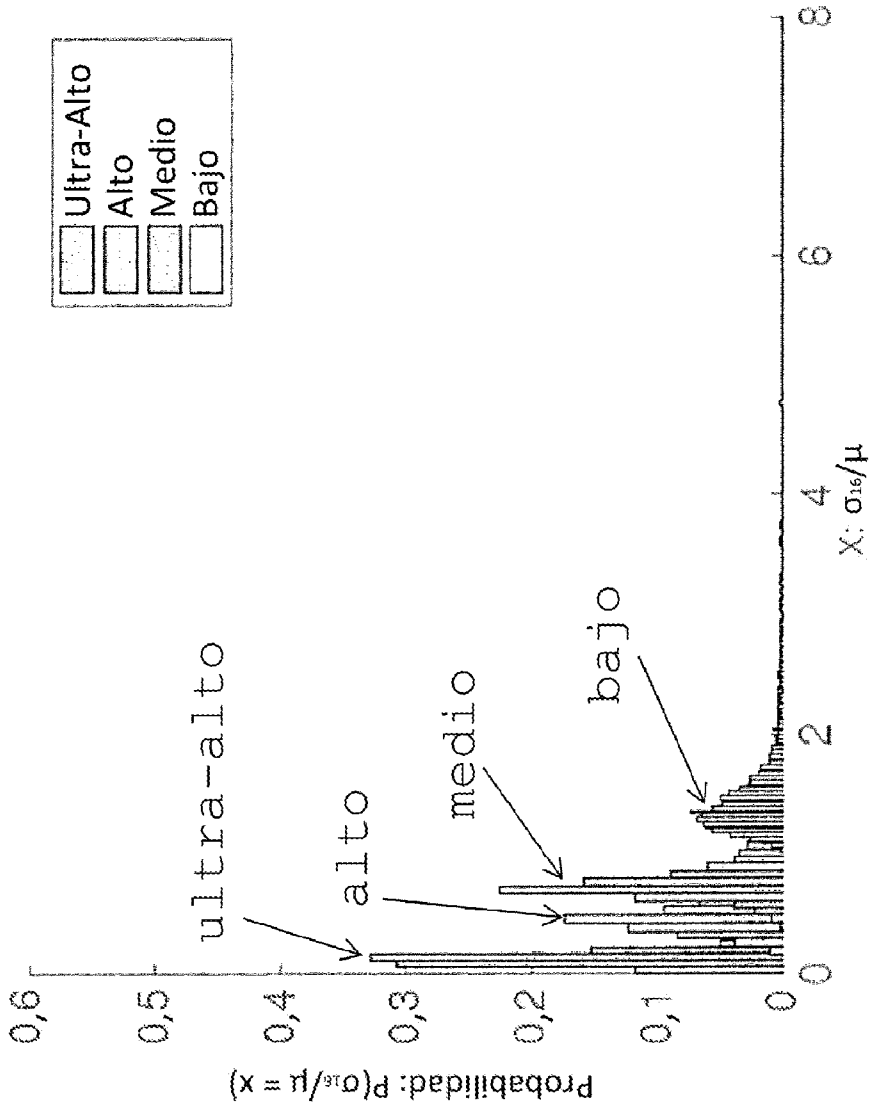

Ultra-Alto
Alto
Medio
Bajo
ultra-alto
alto
medio
bajo
Probabilidad: P(σx/μ = x)
x: σx/μ
(e) Ráfaga a 4 segundo
(f) Ráfaga a 8 segundo

FIGURA 8



(g) Ráfaga a 16 segundo

FIGURA 8

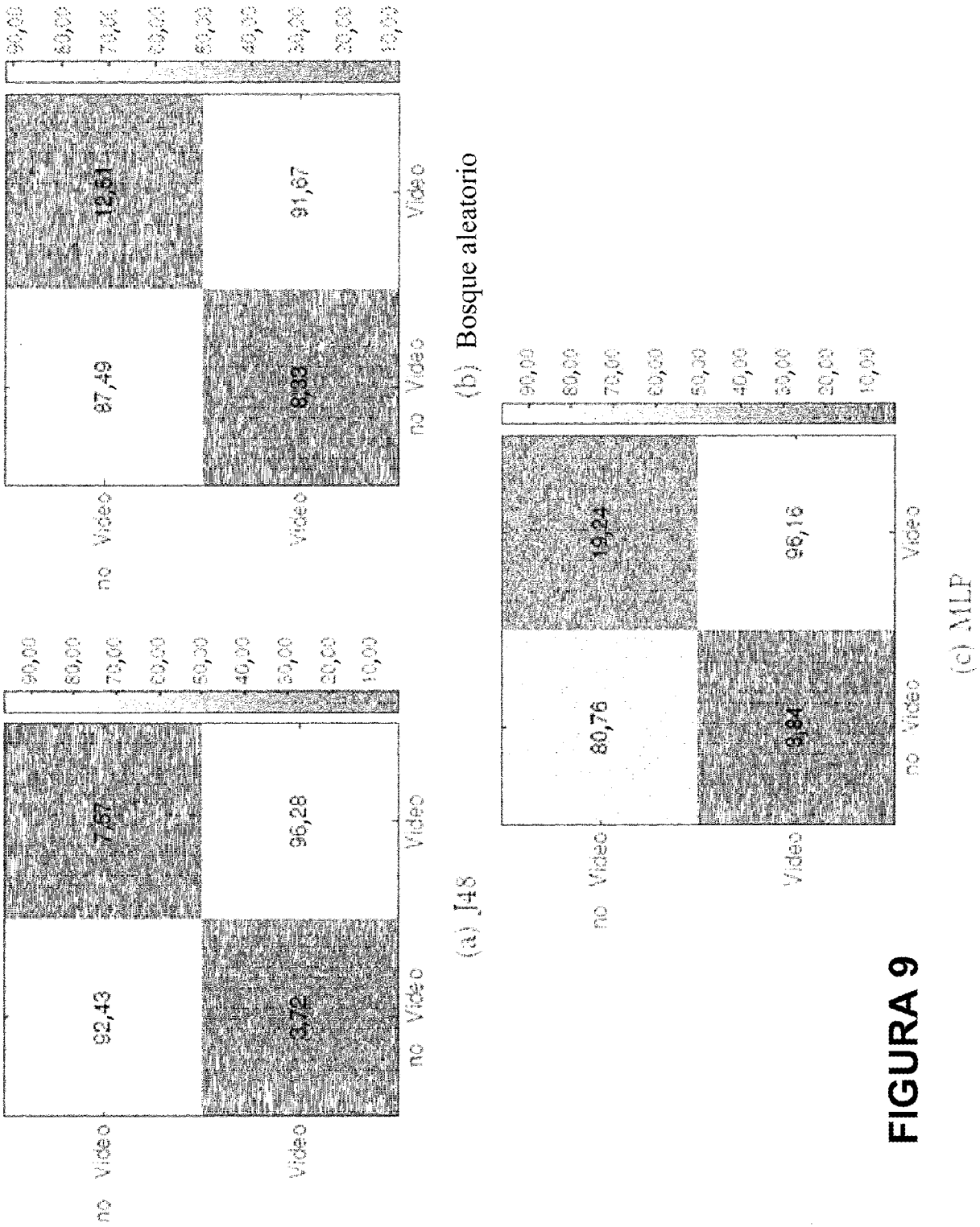

90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
92,43
7,57
3,72
96,28
87,49
12,51
8,33
91,67
80,76
19,24
3,84
96,16
no Video
Video
(a) J48
(b) Bosque aleatorio
(c) MLP

FIGURA 9

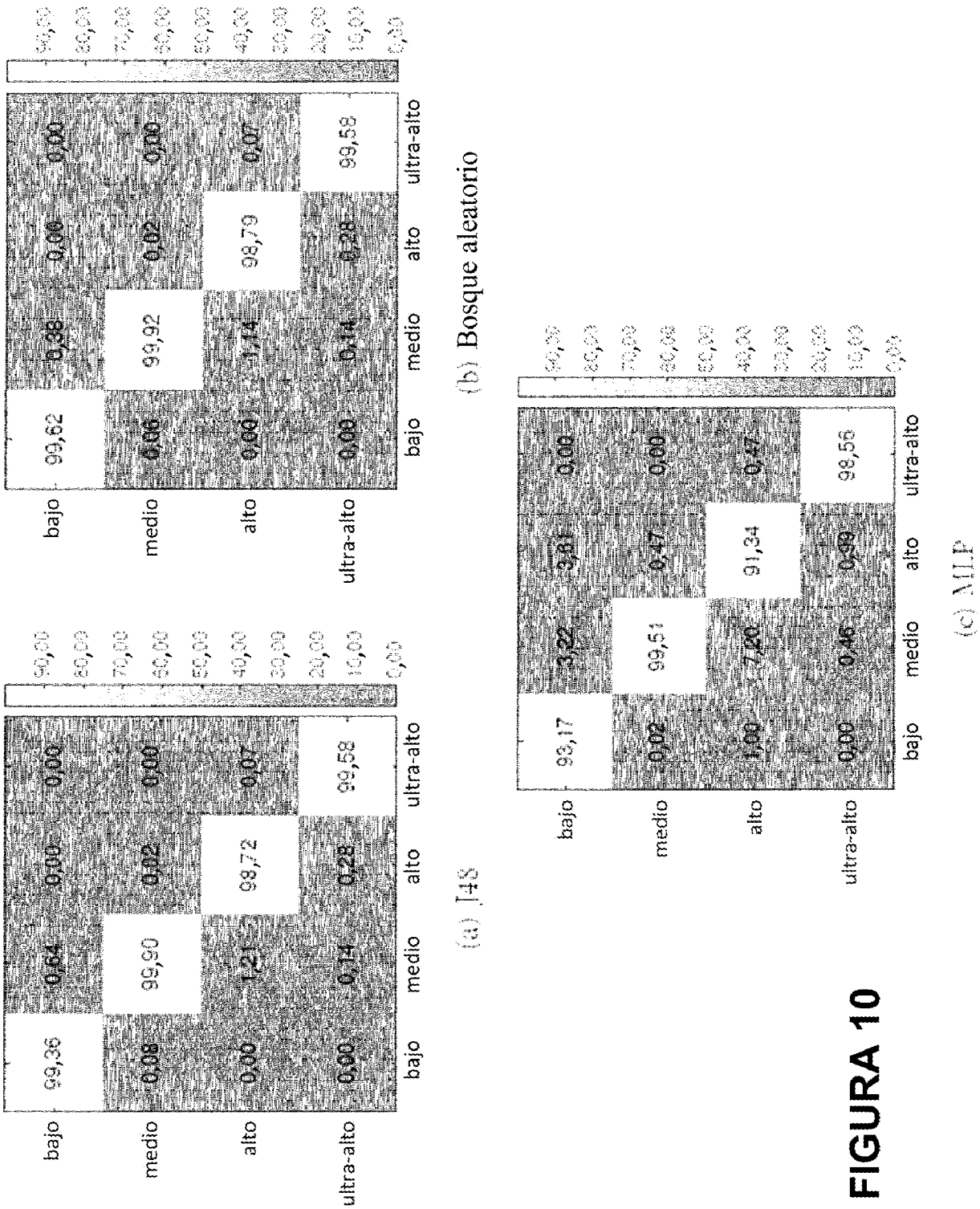
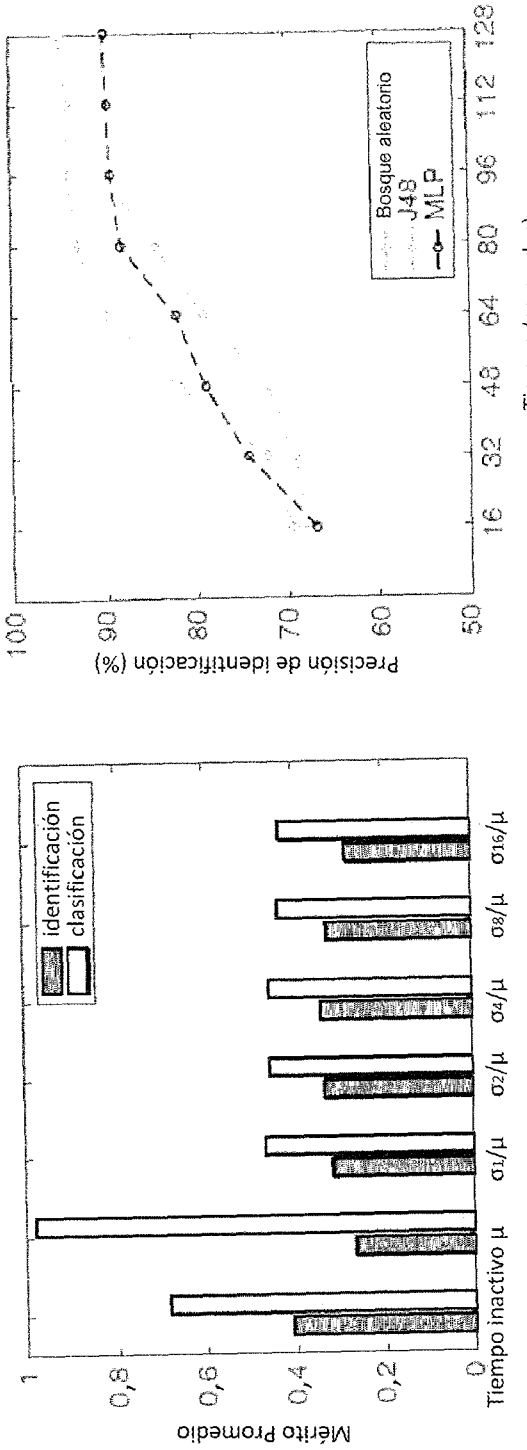
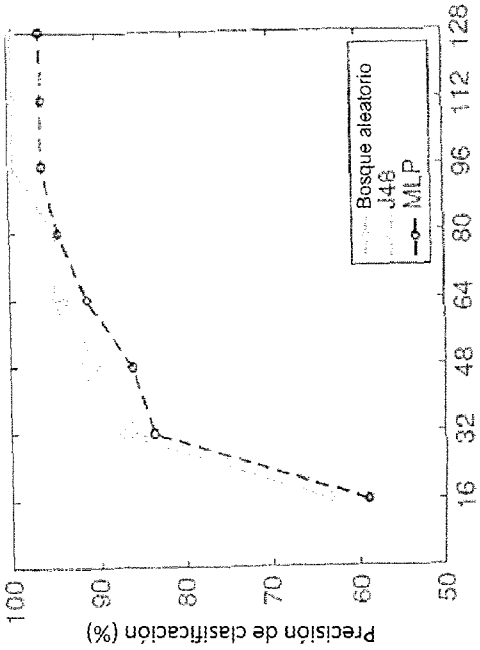

bajo
medio
alto
ultra-alto
99,36
0,64
0,00
0,00
0,08
99,90
0,02
0,00
0,00
1,21
98,72
0,07
0,00
0,14
0,28
99,58
(a) J48
99,62
0,38
0,00
0,00
0,06
99,92
0,02
0,00
0,00
0,14
98,79
0,07
0,00
0,14
0,28
99,58
(b) Bosque aleatorio
93,17
3,22
3,61
0,00
0,02
99,51
0,47
0,00
1,00
7,20
91,34
0,47
0,00
0,46
0,99
98,56
(c) MLP

FIGURA 10



(a) Mérito del atributo

(b) Identificación de video



(c) Clasificación de resolución

FIGURA 11

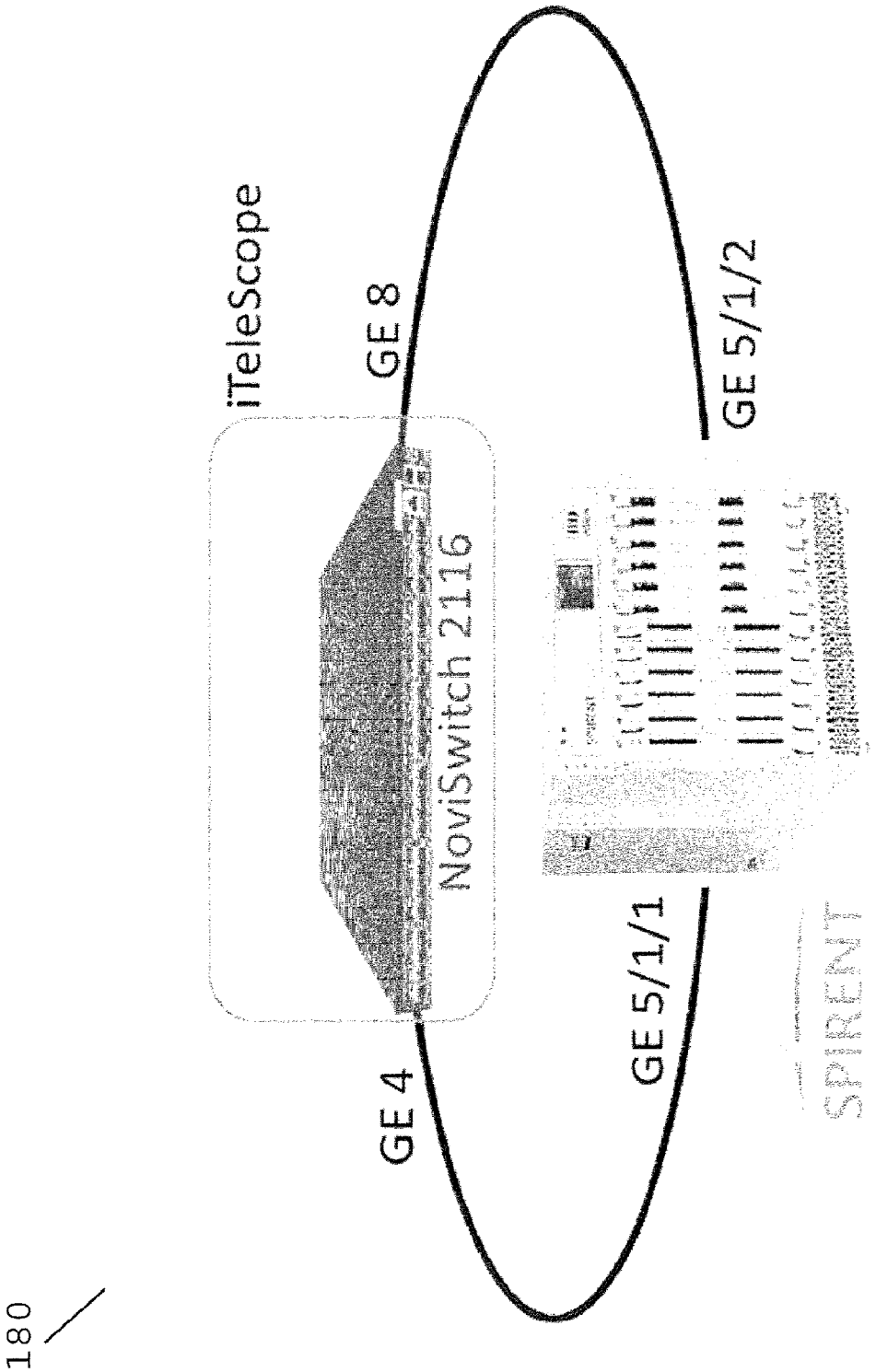

180
iTeleScope
GE 8
GE 5/1/2
NoviSwitch 2116
GE 4
GE 5/1/1
SPIRENT

FIGURA 12

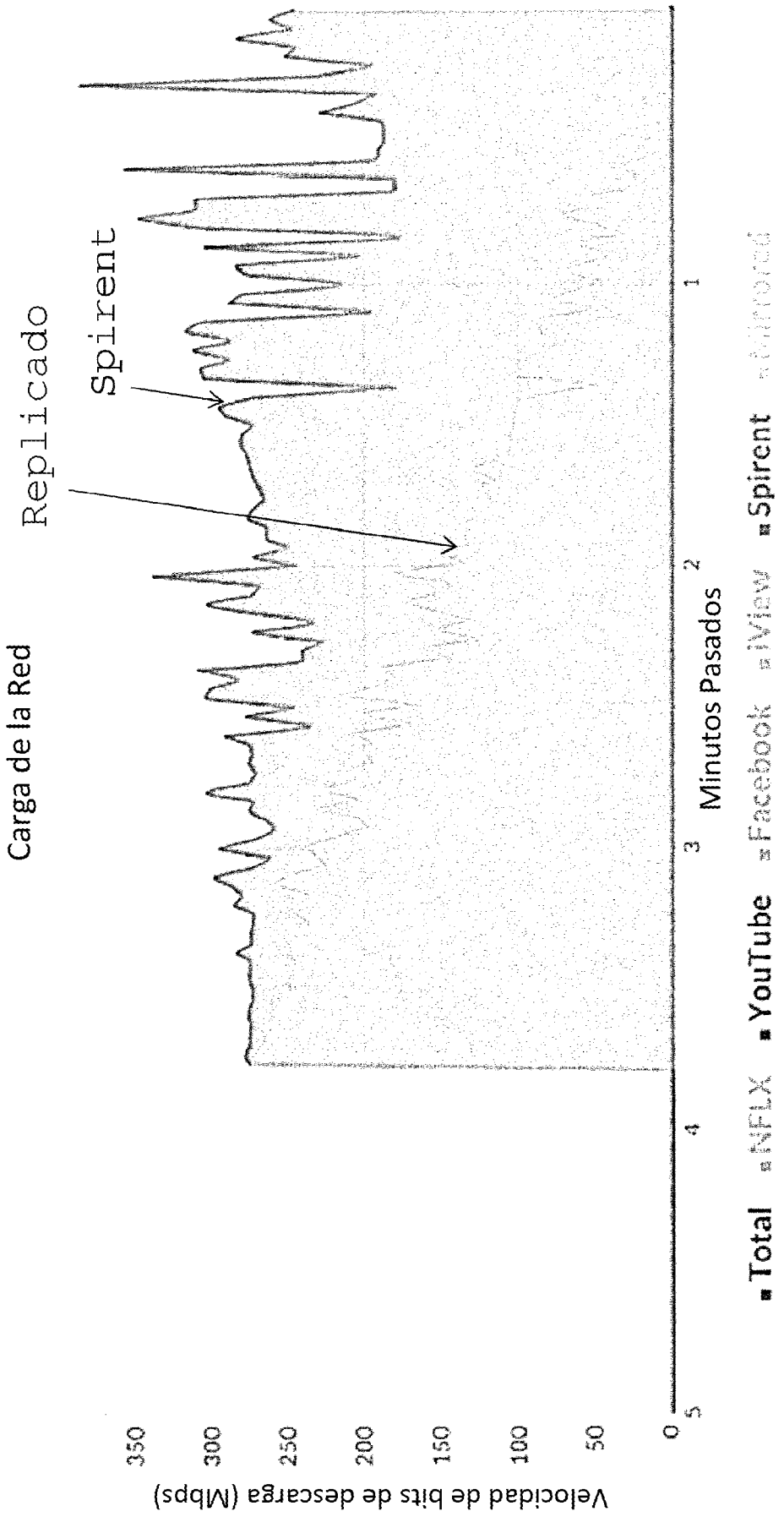

Carga de la Red
Replicado
Spirent
Velocidad de bits de descarga (Mbps)
350
300
250
200
150
100
50
0
1
2
3
4
5
Minutos Pasados
Total
NFLX
YouTube
Facebook
iView
Spirent

FIGURA 13

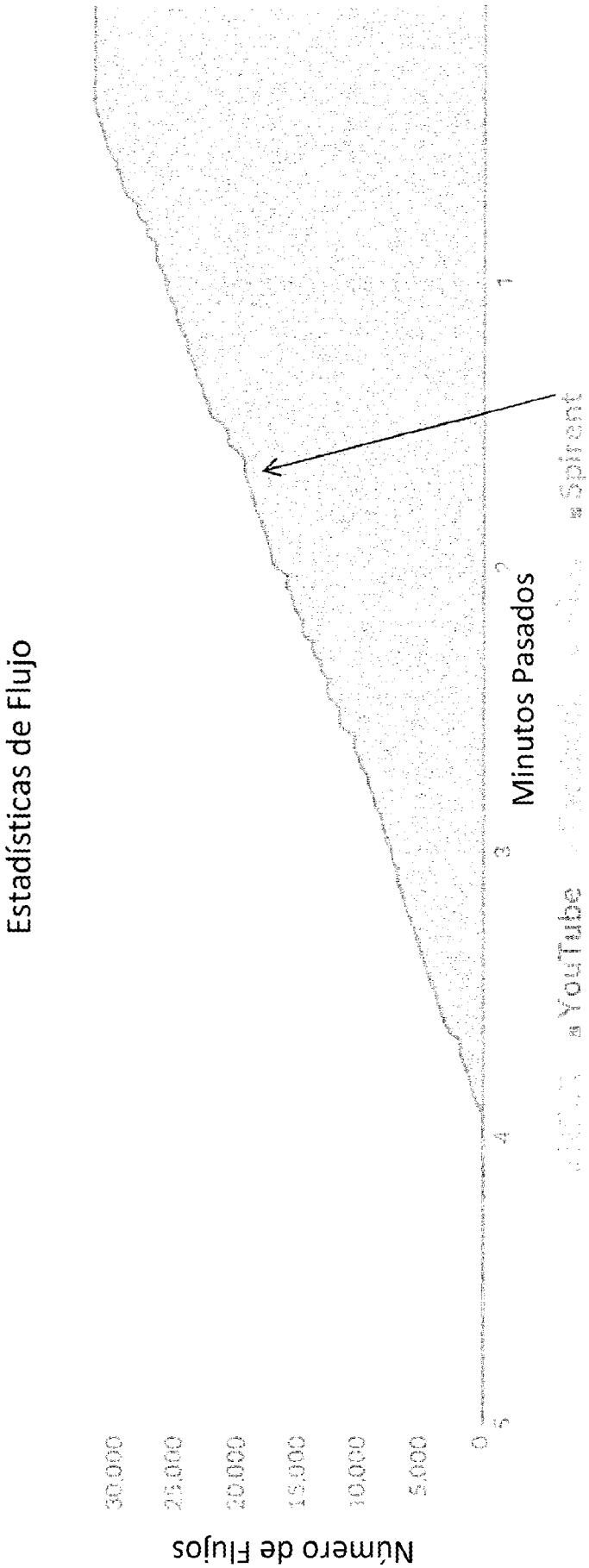
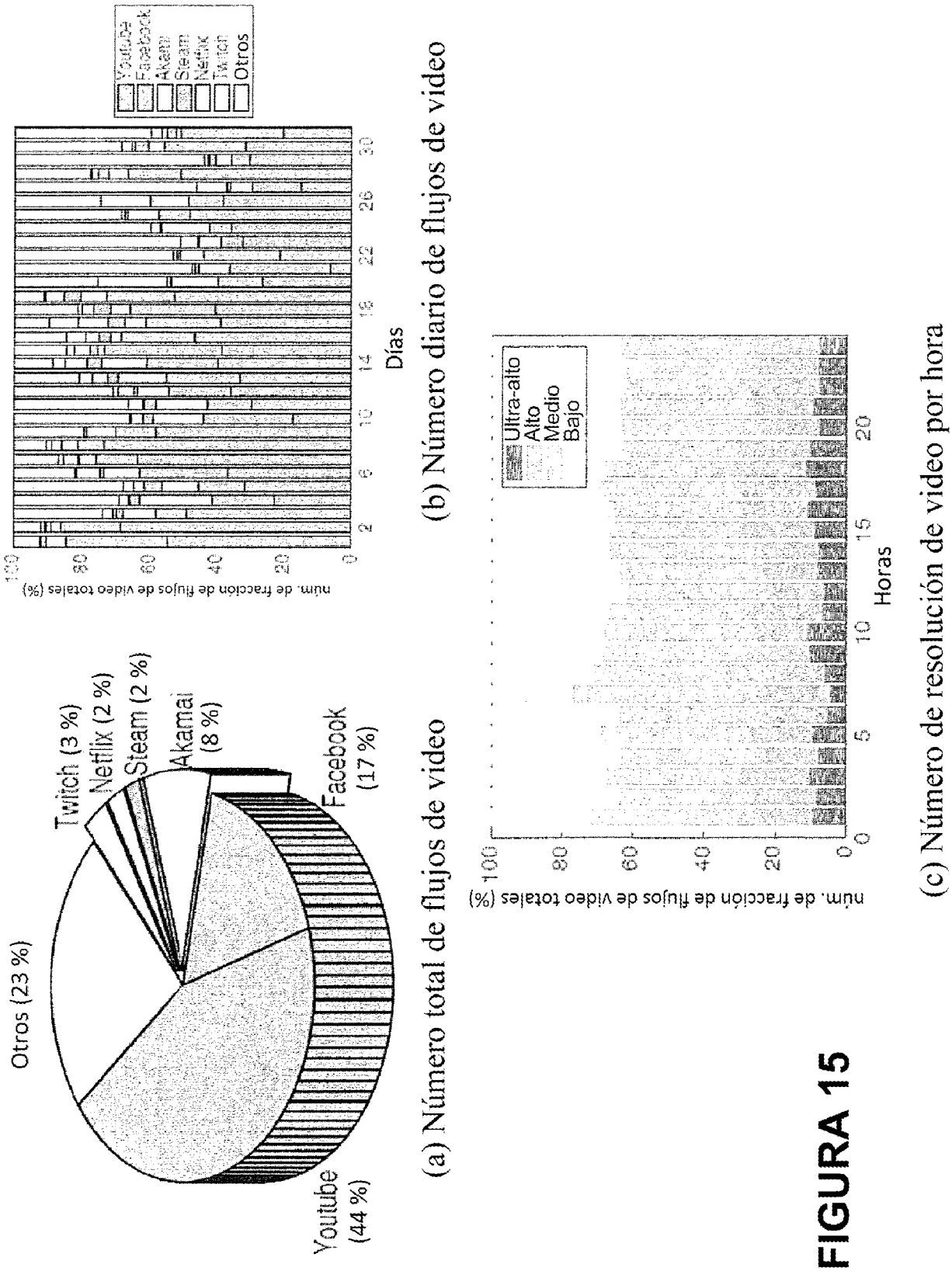

Estadísticas de Flujo
Minutos Pasados
Número de Flujos
30.000
25.000
20.000
15.000
10.000
5.000
0
1
2
3
4
5
YouTube
Sprint

FIGURA 14

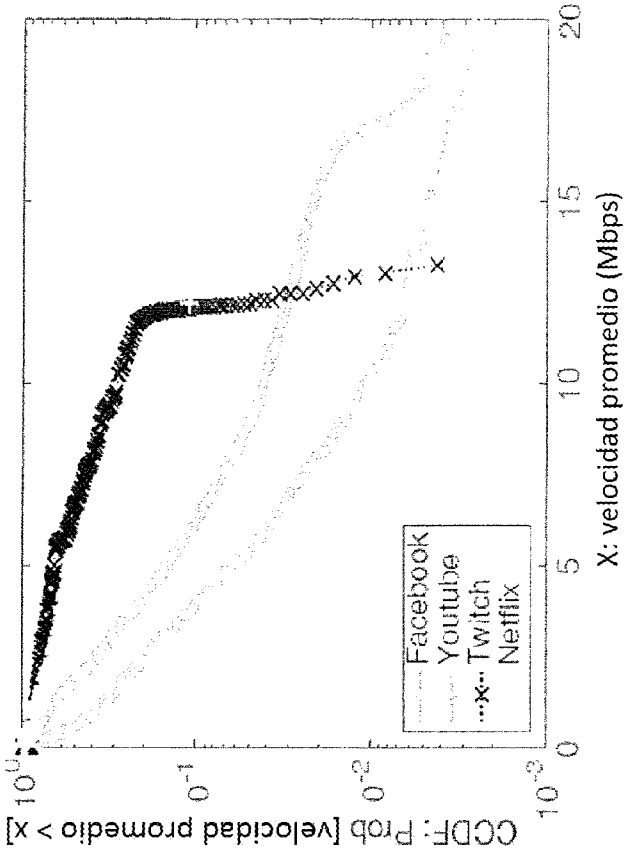


(a) Número total de flujos de video

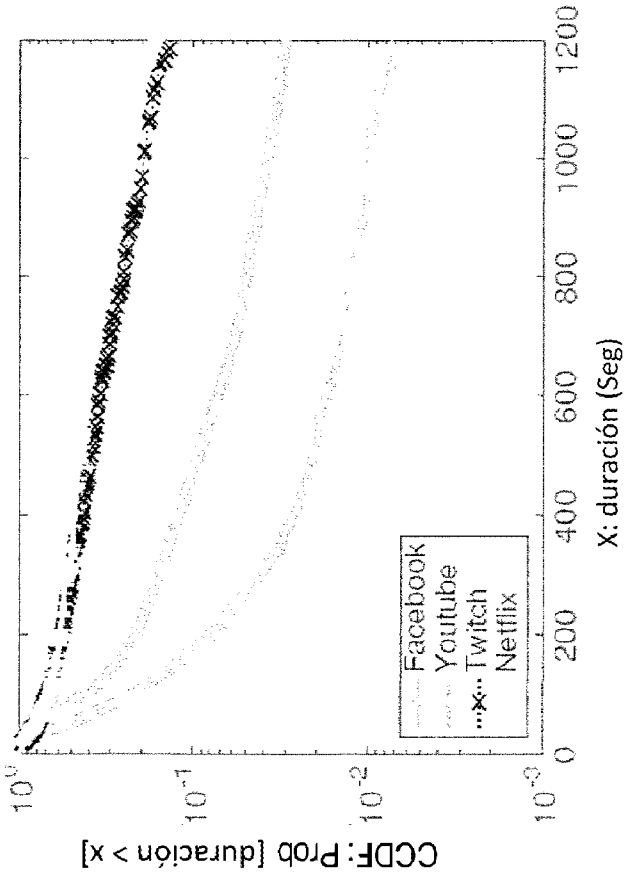
(b) Número diario de flujos de video

(c) Número de resolución de video por hora

FIGURA 15



(b) Velocidad de video



(a) Duración del video

FIGURA 16

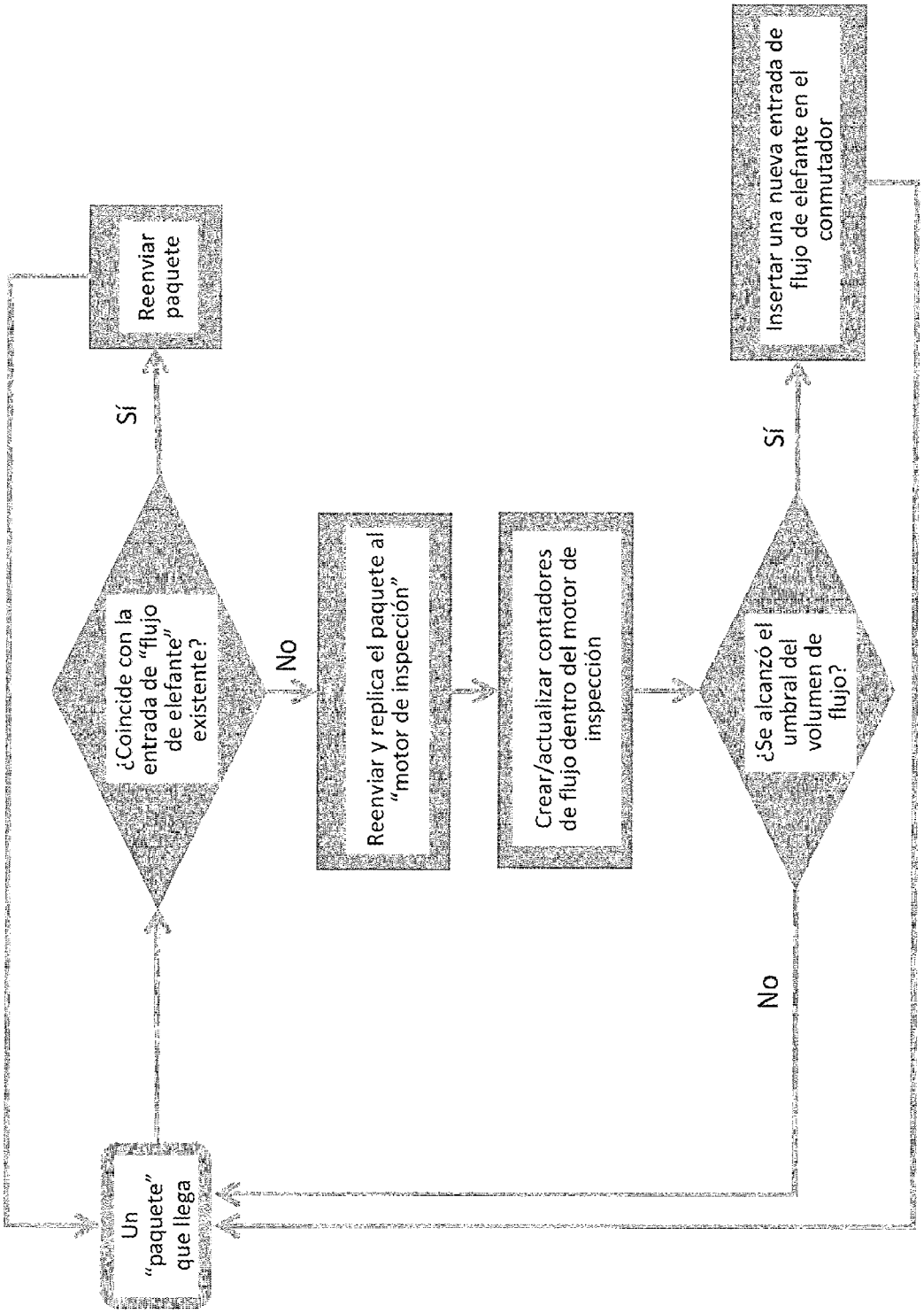

Un "paquete" que llega
¿Coincide con la entrada de "flujo de elefante" existente?
Sí
Reenviar paquete
No
Reenviar y replica el paquete al "motor de inspección"
Crear/actualizar contadores de flujo dentro del motor de inspección
¿Se alcanzó el umbral del volumen de flujo?
Sí
Insertar una nueva entrada de flujo de elefante en el conmutador
No

FIGURA 17

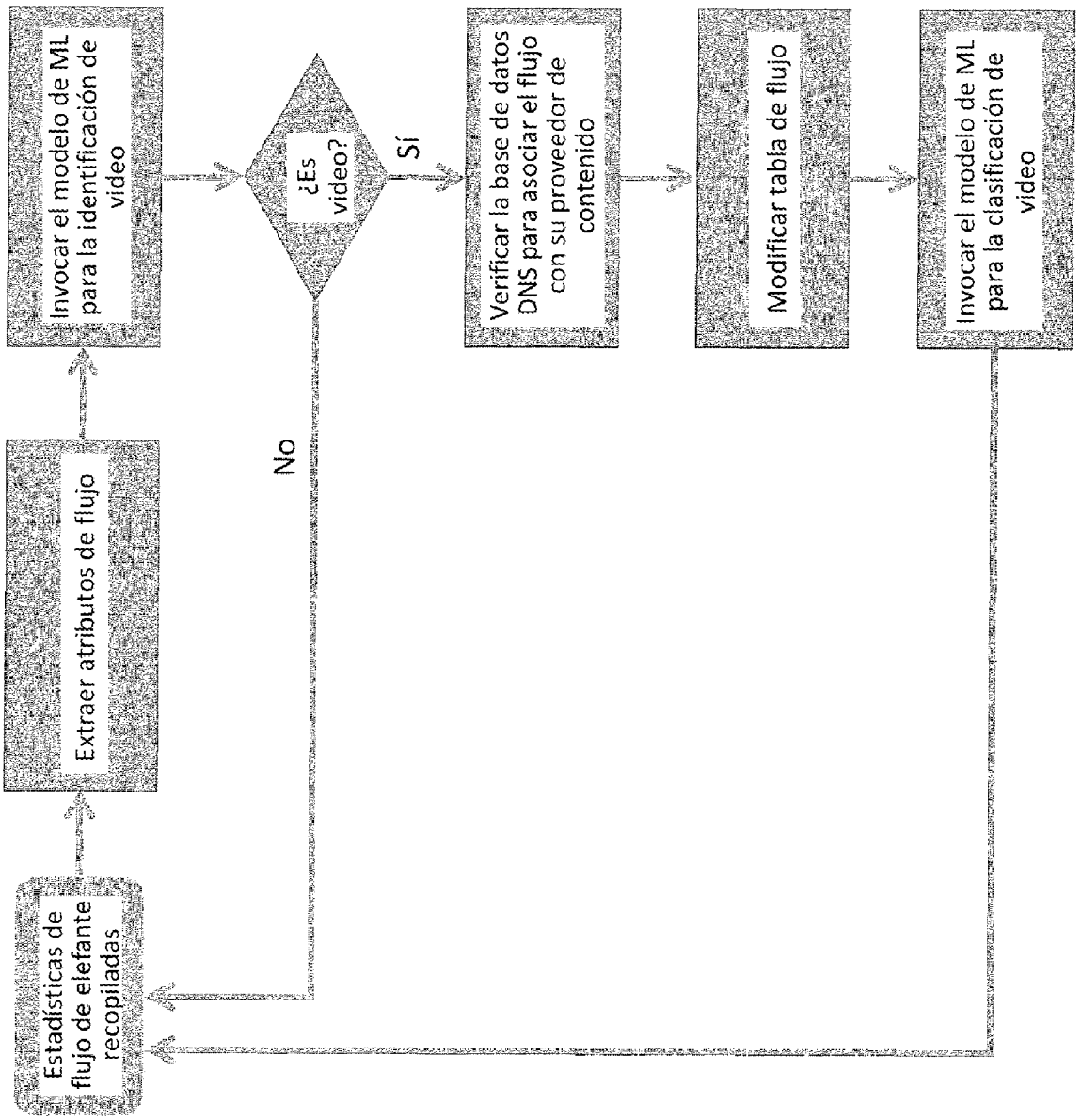

Estadísticas de flujo de elefante recopiladas
Extraer atributos de flujo
Invocar el modelo de ML para la identificación de vídeo
¿Es vídeo?
No
Sí
Verificar la base de datos DNS para asociar el flujo con su proveedor de contenido
Modificar tabla de flujo
Invocar el modelo de ML para la clasificación de vídeo

FIGURA 18