

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 703 153**

51 Int. Cl.:

H04L 29/06 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.02.2015** **E 15382075 (8)**

97 Fecha y número de publicación de la concesión europea: **26.09.2018** **EP 3062474**

54 Título: **Método, dispositivo, producto de programa informático y medio de almacenamiento para distribuir solicitudes de archivos en sistemas de transmisión en continuo adaptables**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
07.03.2019

73 Titular/es:

ALCATEL-LUCENT ESPAÑA, S.A. (100.0%)
C/ María Tubau, 9
28050 Madrid, ES

72 Inventor/es:

RUIZ ALONSO, JAIME;
PÉREZ GARCÍA, PABLO y
VILLEGAS NÚÑEZ, ÁLVARO

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 703 153 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Método, dispositivo, producto de programa informático y medio de almacenamiento para distribuir solicitudes de archivos en sistemas de transmisión en continuo adaptables

5

Antecedentes de la invención

Campo técnico

10 La presente invención se refiere a la entrega de contenido multimedia (por ejemplo, texto, audio, vídeo, software, etc., y cualquier combinación de los mismos) que utilizan un formato de transmisión en continuo adaptable en redes de comunicaciones. Más particularmente, la presente invención se refiere a la evitación de la congestión en servidores de Protocolo de Control de Transmisión (TCP) en sistemas de entrega de transmisión en continuo adaptables.

15

Descripción de la técnica relacionada

Los servidores que proporcionan la entrega de fragmentos de vídeo para Transmisión en Continuo de Vídeo Adaptable sufren, por lo general, de congestión TCP cuando algunas solicitudes se reciben agrupadas en ráfagas. Esto es causado principalmente porque el ancho de banda de red disponible para la entrega de vídeo del servidor al cliente suele estar limitada a un máximo (por la capacidad de ancho de banda de la red o por la latencia de red y colas de búfer del TCP intermedios). Si el servidor es muy rápido en la entrega segmentos de vídeo, la respuesta del servidor al cliente para atender las solicitudes agrupadas en ráfagas provoca la activación de los algoritmos de congestión del TCP (que limitan el rendimiento de entrega del servidor). Los algoritmos de congestión del TCP pueden causar retransmisiones e incluso la pérdida de la conexión, incluso si el ancho de banda de entrega requerido es inferior al disponible.

Diversos algoritmos de congestión del TCP se podrían optimizar para la entrega de vídeo en situaciones particulares, con el fin de controlar la forma en que el tráfico del TCP se puede restablecer o limitar.

30

Muchas variables están involucradas en el rendimiento de salida del TCP, pero tal vez las más representativas son las siguientes:

- BB (ancho de banda de cuello de botella)
- RTT (tiempo de ida y vuelta)
- Enviar y Recibir Búfer Conectores
- RWND del TCP mínima (Ventana De Recepción)
- MTU (Unidad de Transmisión Máxima) de trayectoria

40 A continuación, si se alcanza el ancho de banda disponible, los algoritmos de congestión del TCP causar varios problemas en el servidor:

- Algunos paquetes del TCP podrían retransmitirse, causando más tráfico de red.
- Algunas conexiones del TCP podrían perderse, causando fallos en el servidor a los clientes.
- Las colas del TCP en el servidor crecen lo que requiere más recursos del Servidor para almacenar la información en espera para su envío.
- Se requiere que la CPU del servidor trate con los algoritmos de contención del TCP.

50 Cuando el rendimiento de los servidores es alto (es decir, los servidores son capaces de entregar los archivos solicitados muy rápidamente), tan pronto como las solicitudes de fragmentos de HTTP son recibidas en el lado del servidor (normalmente en el orden de solicitudes 1 Kbyte), los archivos solicitados (fragmentos de vídeo normalmente del orden de 1 Mbyte) se envían a los solicitantes.

55 Sin embargo, ninguno de los algoritmos de congestión del TCP (es decir OSI de nivel 4, Enlace de Transporte en el lado del servidor, TCP) son conscientes del tipo de transporte que se entrega para la entrega de vídeo de transmisión en continuo Adaptable (es decir, OSI de nivel 6, Aplicación, vídeo de transmisión en continuo adaptable del HTTP).

60 Por lo tanto, existe la necesidad de una mejora a nivel de aplicación (OSI de nivel 6) para evitar la congestión del TCP cuando se entrega el vídeo de transmisión en continuo adaptable del HTTP (HAS).

65 El documento WO2013/033565A1 de la técnica anterior describe un sistema de transmisión en continuo del HTTP adaptable, en el que múltiples solicitudes de segmentos pueden estar en mapa o agregarse a un cliente para evitar retrasos del inicio del TCP repetitivos.

Sumario

5 En vista de la presente necesidad de proporcionar servidores con un método para evitar la congestión del TCP en escenarios HAS que supere los inconvenientes antes mencionados, la presente invención proporciona un método y un dispositivo (elemento de red) para la redistribución sin problemas de las solicitudes de archivos multimedia antes de que se reciban por la aplicación del servidor.

10 La presente invención redistribuye las solicitudes teniendo en cuenta uno o más o una combinación de los siguientes parámetros medidos en la red:

- 10 – El tamaño del archivo solicitado, que en algunos sistemas de entrega de Velocidad de Bits Adaptable (ABR) del HTTP puede deducirse de la sintaxis de URL solicitada del archivo
- La ráfaga máxima tolerada que puede soportarse en un período de tiempo
- El ancho de banda máximo de la interfaz de red disponible
- 15 – La retardo máxima tolerada que se puede introducir por el suavizado
- Las direcciones IP de los clientes que solicitan los archivos multimedia: El algoritmo de suavizado propuesto puede aplicarse a las solicitudes relacionadas con ciertos intervalos de direcciones IP y, a continuación, el servidor puede suavizar el flujo aguas abajo de las solicitudes por los intervalos de cliente.

20 El algoritmo de suavizado propuesto se puede implementar en un controlador de un Servidor del TCP o en un dispositivo intermedio que procesa todas las solicitudes de archivos de los clientes.

25 En una realización preferida de la invención, la redistribución de solicitudes se implementa mediante la introducción de un retraso artificial en algunas de las solicitudes recibidas. Además, la solución propuesta puede incluir también un retraso máximo aceptable que se introduce con el fin de garantizar una redistribución suave de las solicitudes en un intervalo de tiempo dado.

30 El retraso propuesto se aplica a un grupo de una o más solicitudes de archivos que cumple con un criterio basado en uno o más de los parámetros antes mencionados para el suavizado.

35 Algunas simplificaciones y omisiones se pueden realizar en el presente sumario, lo que se pretende es destacar e introducir algunos aspectos de las diversas realizaciones a modo de ejemplo, pero no para limitar su alcance. Las descripciones detalladas de las realizaciones a modo de ejemplo preferidas adecuadas para permitir que los expertos normales en la técnica realicen y utilicen los conceptos de invención se continuarán en secciones posteriores.

40 De acuerdo con un aspecto de la invención se proporciona un método para la distribución de solicitudes de archivos que comprenden contenido multimedia para ser entregado en transmisión en continuo del HTTP adaptable, en el que la etapa de distribución comprende la definición de un grupo de solicitudes de todas las solicitudes de archivo recibidas (por ejemplo, en una entrada del servidor de entrega de transmisión en continuo adaptable) en un intervalo de tiempo determinado, el grupo de solicitudes que coinciden en un patrón basado en uno o más parámetros de suavizado, y la aplicación de un retraso (por ejemplo, retraso máximo tolerado) para el grupo definido de solicitudes. Esta etapa se repite cada intervalo de tiempo determinado para la redistribución de las solicitudes de archivos y la entrega del archivo solicitado del contenido multimedia a un cliente por una aplicación de entrega de acuerdo con dicha distribución. La etapa de distribución se realiza antes de recibir las solicitudes de archivos procedentes del cliente por la aplicación de entrega.

50 De acuerdo con otro aspecto de la invención un dispositivo para la distribución de solicitudes de archivos con contenido multimedia (por ejemplo, vídeo), que es un elemento de red conectado a un sistema de entrega de transmisión en continuo adaptable del HTTP, se proporciona, comprendiendo medios para implementar el método descrito anteriormente. El dispositivo propuesto puede ser un servidor (por ejemplo, un controlador cargable en el servidor implementa el algoritmo de evitación de la congestión propuesta) o un dispositivo intermedio.

55 De acuerdo con un último aspecto de la invención, se proporciona un producto de programa informático, que comprende instrucciones ejecutables por ordenador para realizar cualquiera de las etapas del método divulgado anteriormente, cuando el programa se ejecuta en un ordenador y un medio de almacenamiento de datos digitales se proporciona también codificando un programa ejecutable por máquina de instrucciones para realizar cualquiera de las etapas del método anteriormente divulgado.

60 Estos y otros aspectos de la invención serán evidentes a partir de y se aclararán con referencia a las realizaciones descritas a continuación.

Las realizaciones presentadas tienen potencialmente las siguientes ventajas en comparación con la técnica anterior:

- 65 – El método propuesto proporciona servidores más robustos que los servidores que no implementan el método

propuesto para evitar la congestión y permite un mejor servicio en el despliegue real de Servicios de Entrega de Transmisión en Continuo de Vídeo masivos.

- El método propuesto mejora el despliegue de servicios masivos donde las solicitudes de ráfaga ocasionales se generan aleatoriamente por muchas razones diferentes (por ejemplo, dispositivos de red, enrutadores de red, firewalls de red, despliegues OTT,...).

Breve descripción de los dibujos

Algunas realizaciones del método, sistema y dispositivo de acuerdo con las realizaciones de la presente invención se describen a continuación, a modo de ejemplo solamente, y con referencia a los dibujos adjuntos, en los que:

La Figura 1 muestra una representación gráfica de la distribución del número de solicitudes de archivo a lo largo del tiempo, en un posible escenario de red utilizando un método de distribución de solicitudes conocido en el estado de la técnica.

La Figura 2 muestra una representación gráfica del ancho de banda medio del TCP en función del tiempo, en el escenario y método de distribución de solicitudes de la Figura 1.

La Figura 3 muestra una representación gráfica de un ancho de banda máximo del TCP en función del tiempo, en el escenario y método de distribución de solicitudes de la Figura 1.

La Figura 4 muestra una representación gráfica del ancho de banda mínimo del TCP en función del tiempo, en el escenario y método de distribución de solicitudes de la Figura 1.

La Figura 5 muestra una representación gráfica del tiempo medio de respuesta y descarga de un fragmento de archivo para la entrega en transmisión en continuo adaptable, en el escenario y método de distribución de solicitudes de la Figura 1.

La Figura 6 muestra una representación gráfica de la distribución del número de solicitudes de archivo a lo largo del tiempo, en un posible escenario de red utilizando un método de redistribución de suavizado para las solicitudes de acuerdo con una realización preferida de la presente invención.

La Figura 7 muestra una representación gráfica de un ancho de banda máximo del TCP en función del tiempo, en el escenario y método de distribución de solicitudes de la Figura 6 de acuerdo con una realización preferida de la presente invención.

La Figura 8 muestra una representación gráfica del ancho de banda medio del TCP en función del tiempo, en el escenario y método de distribución de solicitudes de la Figura 6 de acuerdo con una realización preferida de la presente invención.

La Figura 9 muestra una representación gráfica del ancho de banda mínimo del TCP en función del tiempo, en el escenario y método de distribución de solicitudes de la Figura 6 de acuerdo con una realización preferida de la presente invención.

En todas las figuras los mismos números de referencia se refieren a elementos similares.

Descripción de las realizaciones

Las presentes invenciones pueden realizarse en otros dispositivos, sistema y/o métodos específicos. Las realizaciones descritas han de considerarse en todos los aspectos solo como ilustrativas y no restrictivas. En particular, el alcance de la invención se indica por las reivindicaciones adjuntas más que por la descripción y las figuras de la presente memoria. Todos los cambios que caigan dentro del significado e intervalo de equivalencia de las reivindicaciones han de incluirse dentro de su alcance.

En una posible realización de la invención, el siguiente algoritmo de suavizado puede implementarse en un controlador de red de TCP:

- Distribuir suavemente cada solicitud de Fragmento de Vídeo Adaptable recibida, por ejemplo, que se filtra por la sintaxis URL a GET *.ts, mediante la introducción de un retraso máximo de 40 milisegundos.
- Identificar dos grupos de solicitudes: un primer grupo para las solicitudes de los clientes en los intervalos de direcciones IP de 192.168.12.0 a 192.168.12.255, y un segundo grupo para las solicitudes de los clientes en los rangos de direcciones IP de 192.168.13.0 a 192.168.13.255.

En esta realización de la invención, el algoritmo en el controlador del TCP es transparente a todas las solicitudes HTTP GET excepto a las solicitudes GET *.ts. Todas las solicitudes recibidas con este patrón se retrasan a

continuación 40 milisegundos, y, después de 40 milisegundos, un planificador envía todas las solicitudes recibidas perfectamente redistribuidos en los próximos 40 milisegundos, con un retraso constante entre una y la siguiente solicitud. Esta redistribución de las solicitudes puede recalcularse cada milisegundo.

5 En otra realización posible de la invención, el siguiente algoritmo de suavizado puede implementarse en un controlador de red TCP:

- Distribuir suavemente la solicitud de Fragmento de Vídeo Adaptable recibida, por ejemplo, que se filtra por la sintaxis URL a `GET/QualityLevels($number)/Fragments(video=*)`, mediante la introducción de un retraso máximo de 40 milisegundos, de acuerdo a la tasa de bits $\$number$ del fragmento solicitado.

10 En esta realización alternativa de la invención, el algoritmo en el controlador del TCP es transparente a todas las solicitudes GET HTTP, excepto a las solicitudes `GET/QualityLevels($number)/Fragments(video=*)`. Todas las solicitudes recibidas con este patrón se retrasan 40 milisegundos y, después de 40 milisegundos, un planificador envía todas las solicitudes recibidas perfectamente redistribuidas en los próximos 40 milisegundos. El retraso entre una solicitud, $solicitud_{-1}$, y la siguiente solicitud consecutiva, $solicitud_{-2}$, es proporcional a la tasa de bits $\$number$ e indirectamente proporcional a la suma de todos los valores de este campo $\$number$ en la ventana de 40 milisegundos. El retraso después de la $solicitud_{-1}$, es igual a $[40 \text{ ms} * ((\$number \text{ de } solicitud_{-1}) / (\text{suma de todos los } \$number \text{ de solicitudes recibidas en los 40 milisegundos}))]$. Esta redistribución de las solicitudes puede recalcularse cada milisegundo.

15 La Figura 1 muestra un ejemplo de un escenario para un caso de prueba real, en el que 3000 clientes simulados de transmisión en continuo Adaptable HLS que solicitan fragmentos HLS, representan el número de solicitudes de archivos por milisegundo (ms). El N°. de solicitudes se representa en el eje vertical y el tiempo (solo un detalle de un segundo intervalo) se representa en el eje horizontal. Las solicitudes se agrupan en pequeñas ráfagas de cada 40 ms.

20 En el escenario propuesto de la Figura 1, se requiere una media de 6,5 Gbits por segundo a través de una interfaz de 10 Gbps Ethernet como se muestra en la Figura 2, que representa el ancho de banda de salida medio del servidor (en Megabits por segundo) en función del tiempo. En este escenario, la ráfaga en las solicitudes de los clientes, alcanzando picos de más de 10 solicitudes en el mismo milisegundo, hacen que la interfaz alcance picos de ancho de banda de TCP inmediatos del ancho de banda máximo disponible, casi hasta 10 Gbps, como se muestra en la Figura 3. Estos picos causan la activación de los algoritmos de congestión del TCP y las ráfagas hacen que el ancho de banda Mínimo del TCP medido a intervalos instantáneos, mostrado en la Figura 4, sea muy bajo en comparación con el ancho de banda medio de la Figura 2.

25 La Figura 5 muestra que el tiempo medio de entrega de un fragmento de archivo (segmento multimedia) durante esta prueba, utilizando la entrega en transmisión en continuo adaptable, sea de aproximadamente 20 ms.

30 De acuerdo con una realización de la invención, con la finalidad de evitar las ráfagas mencionadas en la red que causan la congestión del TCP a la salida del servidor, en lugar de recibir las solicitudes agrupadas como en la Figura 1, el servidor recibe las solicitudes tal como se muestra en la Figura 6: las solicitudes recibidas en la ráfaga se redistribuyen mediante la introducción de un retraso máximo en algunas de las mismas. Por lo tanto, los anchos de banda máximo y mínimo instantáneos del TCP presentes, mostrados en las Figuras 7 y 9, respectivamente, son muy similares al ancho de banda medio del TCP que se muestra en la Figura 8.

35 Un experto en la materia reconocerá fácilmente que las etapas de los diversos métodos descritos anteriormente se pueden realizar por ordenadores programados. En la presente memoria, algunas realizaciones pretenden cubrir también los dispositivos de almacenamiento de programas, por ejemplo, medios de almacenamiento de datos digitales, que son legibles por máquina o por ordenador y codificar programas de instrucciones ejecutables por máquina o ejecutables por ordenador, en los que dichas instrucciones realizan todas o algunas de las etapas de dichos métodos descritos anteriormente. Los dispositivos de almacenamiento de programas pueden, por ejemplo, ser memorias digitales, medios de almacenamiento magnéticos, tales como cintas magnéticas y discos magnéticos, discos duros, o medios de almacenamiento de datos digitales ópticamente legibles. Las realizaciones pretenden también cubrir los ordenadores programados para realizar dichas etapas de los métodos descritos anteriormente.

40 La descripción y los dibujos ilustran meramente los principios de la invención. La invención se define por el alcance de las reivindicaciones independientes.

45 Además, todos los ejemplos citados en la presente memoria pretenden tener principalmente expresamente solo fines pedagógicos para ayudar al lector en la comprensión de los principios de la invención y los conceptos aportados por el uno o más inventores para la promoción de la técnica, y se deben interpretar como no estando limitados a tales ejemplos y condiciones específicamente mencionadas.

50 Se debe apreciar por los expertos en la materia que cualquier diagrama de bloques en la presente memoria representa vistas conceptuales de circuitos ilustrativos que incorporan los principios de la invención. Del mismo

modo, se apreciará que cualquier gráfico de flujo, diagramas de flujo, diagramas de transición de estados, pseudo código, y similares representan diversos procesos que pueden estar sustancialmente representados en un medio legible por ordenador y ejecutado así por un ordenador o procesador, independientemente de si dicho ordenador o procesador se muestra explícitamente o no.

REIVINDICACIONES

1. Un método para distribuir solicitudes de archivo de contenido multimedia a entregarse en transmisión en continuo adaptable, que comprende:

5 - recibir solicitudes de archivo a partir de al menos un cliente y realizar una distribución de las solicitudes de archivo recibidas para entregar el archivo solicitado de contenido multimedia al cliente mediante una aplicación de entrega de acuerdo con la distribución,
caracterizado por que la etapa de distribución de las solicitudes de archivos recibidas se realiza antes de que
10 las solicitudes de archivos sean recibidos por la aplicación de entrega y consiste en:

15 - filtrar un grupo de solicitudes de las solicitudes de archivo recibidas en un intervalo de tiempo determinado basándose en al menos un parámetro de suavizado que se selecciona de: un tamaño estimado del archivo solicitado, una ráfaga máxima tolerada en un período de tiempo, un ancho de banda de interfaz de red máximo disponible para entregar el archivo solicitado, un retraso máximo tolerado para suavizar el archivo solicitado y un rango de direcciones IP del cliente;
- aplicar un retraso, que es constante o dependiente de la tasa de bits, al grupo de filtrado de solicitudes, para distribuir dicho grupo filtrado de solicitudes en el tiempo;

20 y el método comprende además repetir la etapa de distribución cada intervalo de tiempo determinado.

2. Un elemento de red de un sistema de de transmisión en continuo adaptable para la distribución de las solicitudes de archivo de contenido multimedia, que comprende:

25 - medios para recibir las solicitudes de archivos de al menos un cliente

caracterizado por que comprende además:

30 - una aplicación de entrega para entregar un archivo solicitado de contenido multimedia a un cliente,
- medios para distribuir las solicitudes de archivo recibidas configuradas para realizar una distribución de las solicitudes de archivo recibidas antes de que las solicitudes de archivos sean recibidas por la aplicación de entrega y medios para distribuir las solicitudes de archivo recibidas configurados para:
- filtrar un grupo de solicitudes procedentes de las solicitudes de archivo recibidas en un intervalo de tiempo determinado basándose en al menos un parámetro de suavizado que se selecciona de: un tamaño estimado del
35 archivo solicitado, una ráfaga máxima tolerada en un período de tiempo, un ancho de banda de interfaz de red máximo disponible para entregar el archivo solicitado, un retraso máximo tolerado para suavizar el archivo solicitado y un rango de direcciones IP del cliente;
- aplicar un retraso, que es constante o dependiente de la tasa de bits, al grupo de filtrado de solicitudes, para distribuir dicho grupo filtrado de solicitudes en el tiempo;
40 - repetir la etapa de distribución cada intervalo de tiempo determinado.

3. El elemento de red de acuerdo con la reivindicación 2, que es un servidor de de transmisión en continuo adaptable del HTTP.

45 4. El elemento de red según la reivindicación 2, que es un nodo intermedio conectado a un servidor de transmisión en continuo adaptable.

50 5. Un producto de programa informático que comprende instrucciones ejecutables por ordenador para realizar el método de acuerdo con la reivindicación 1, cuando el programa se ejecuta en un ordenador.

6. Un medio de almacenamiento de datos digitales que codifica un programa de instrucciones ejecutable por máquina para realizar un método de acuerdo con la reivindicación 1.

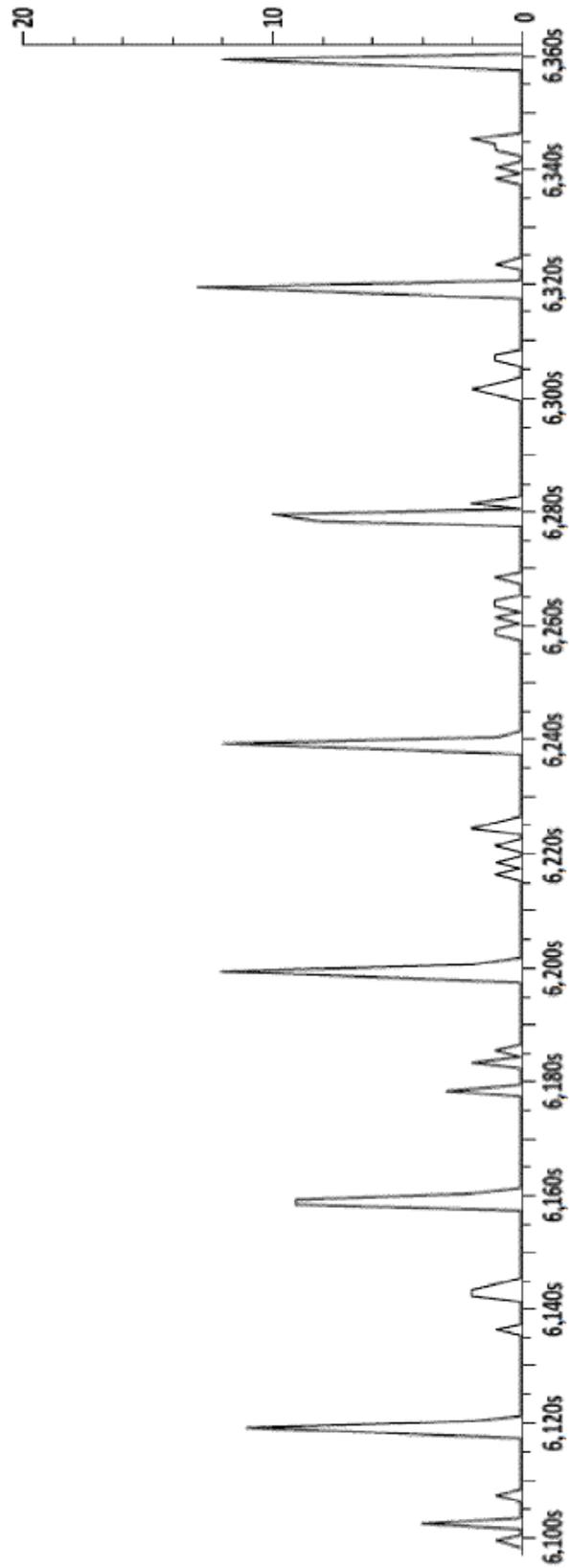


FIG. 1
TÉCNICA ANTERIOR

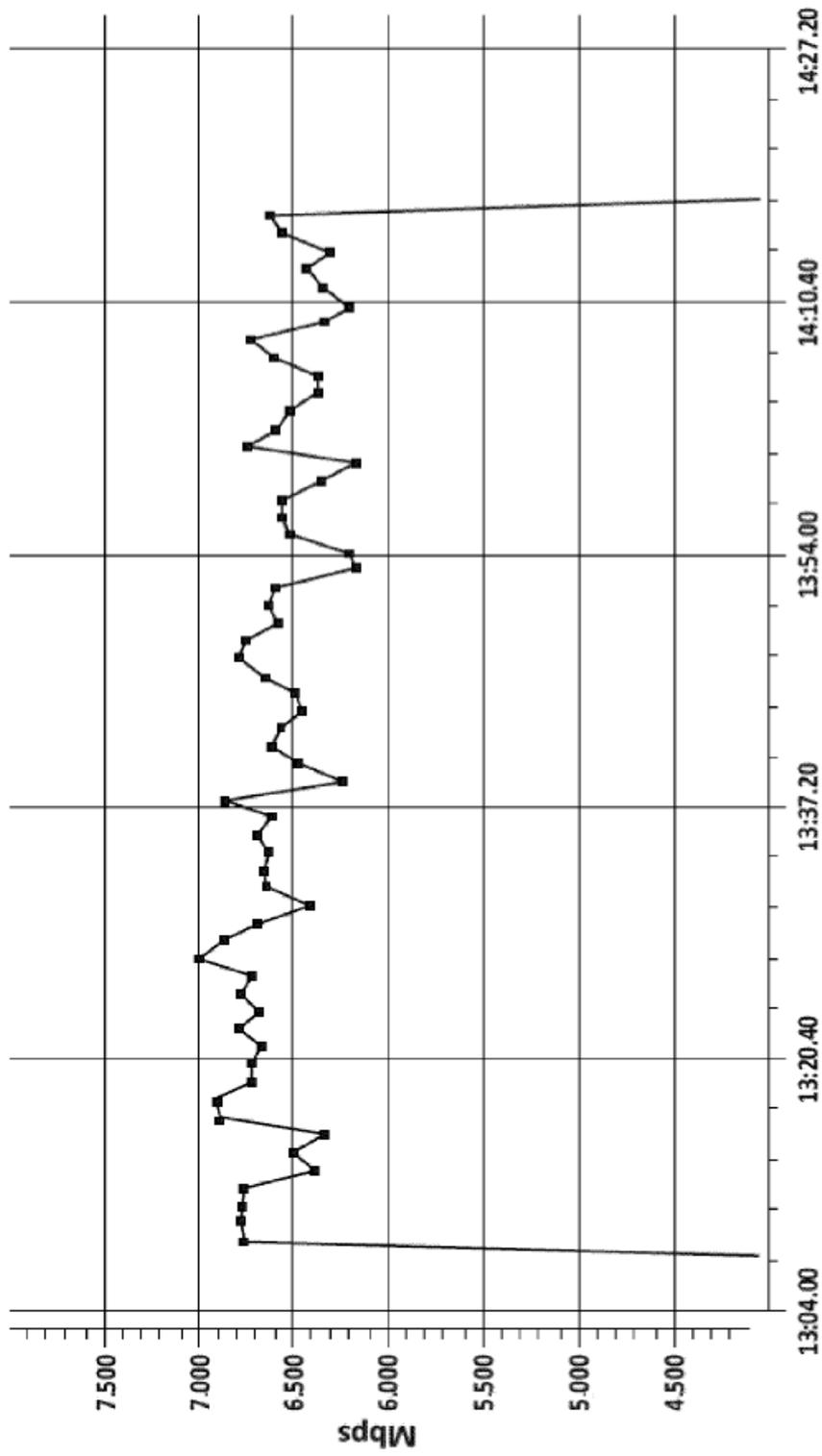


FIG. 2
TÉCNICA ANTERIOR

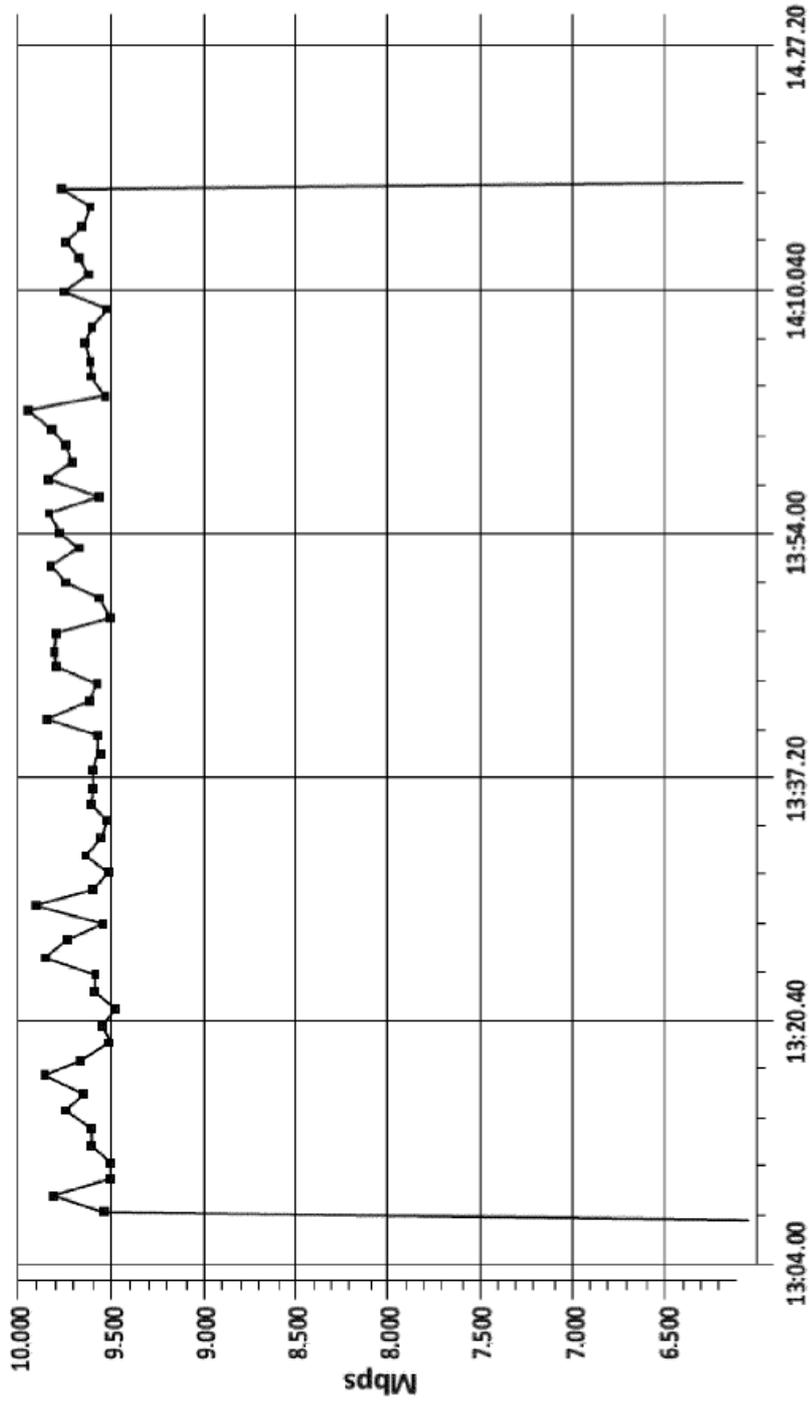


FIG. 3

TÉCNICA ANTERIOR

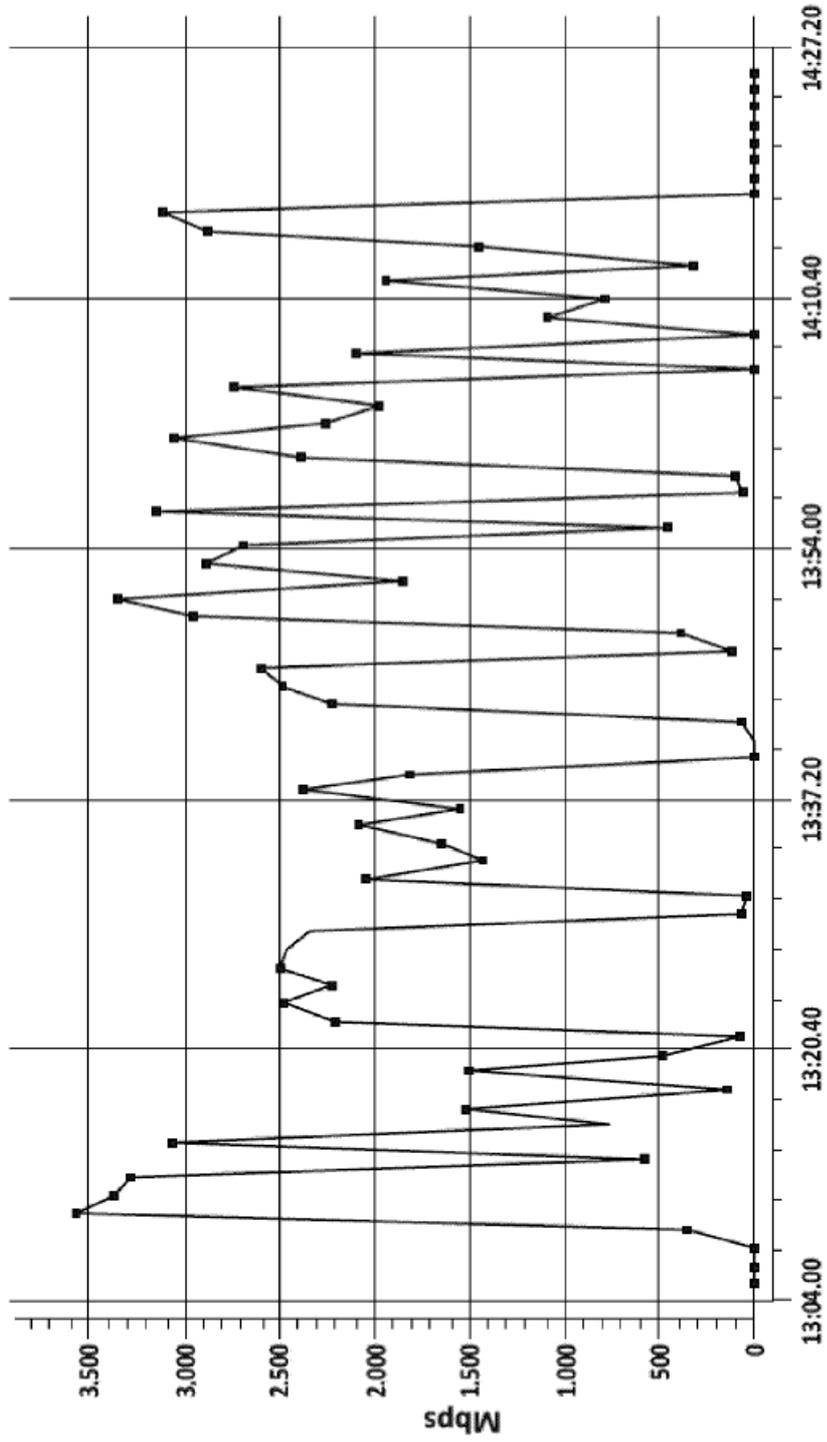


FIG. 4

TÉCNICA ANTERIOR

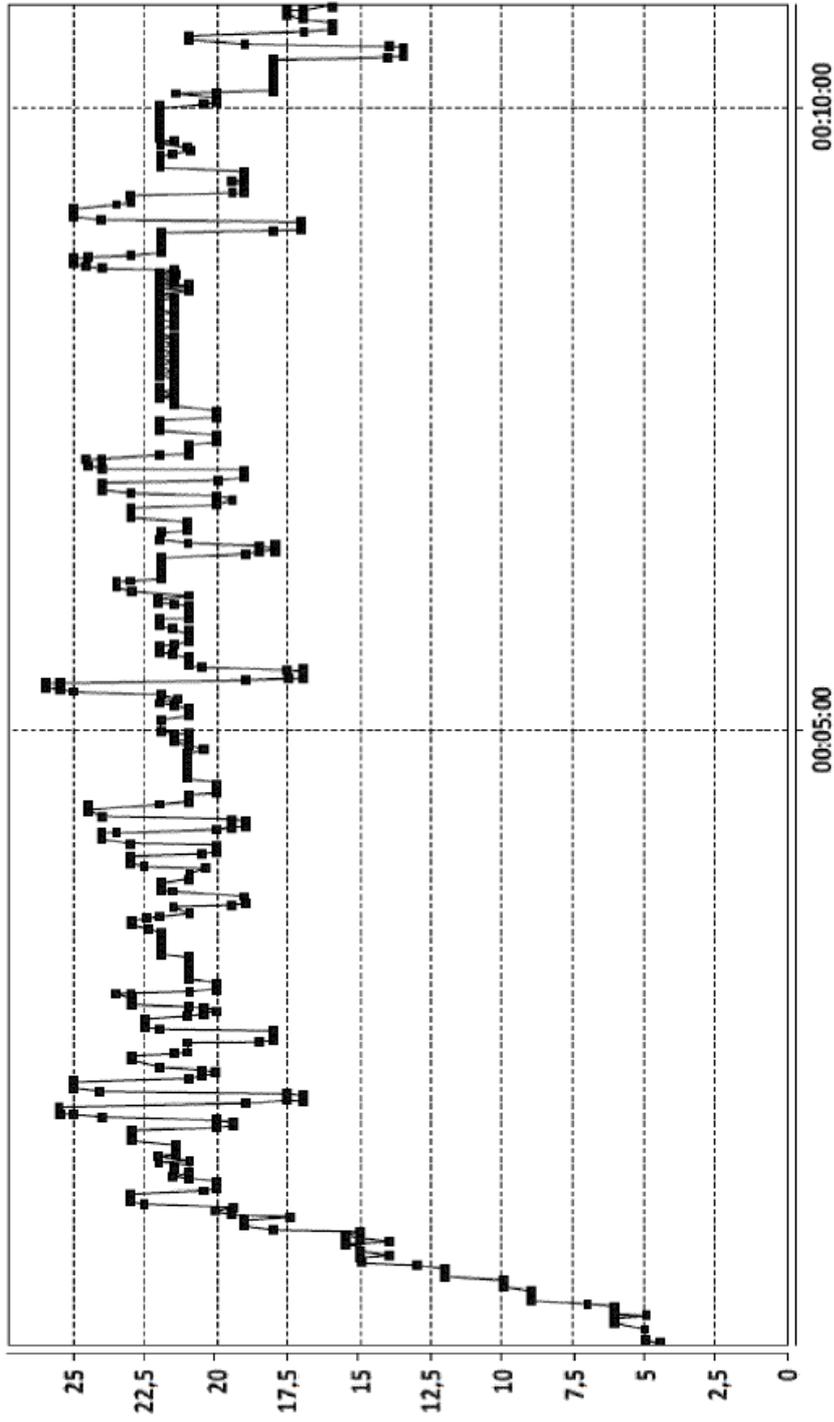


FIG. 5

TÉCNICA ANTERIOR

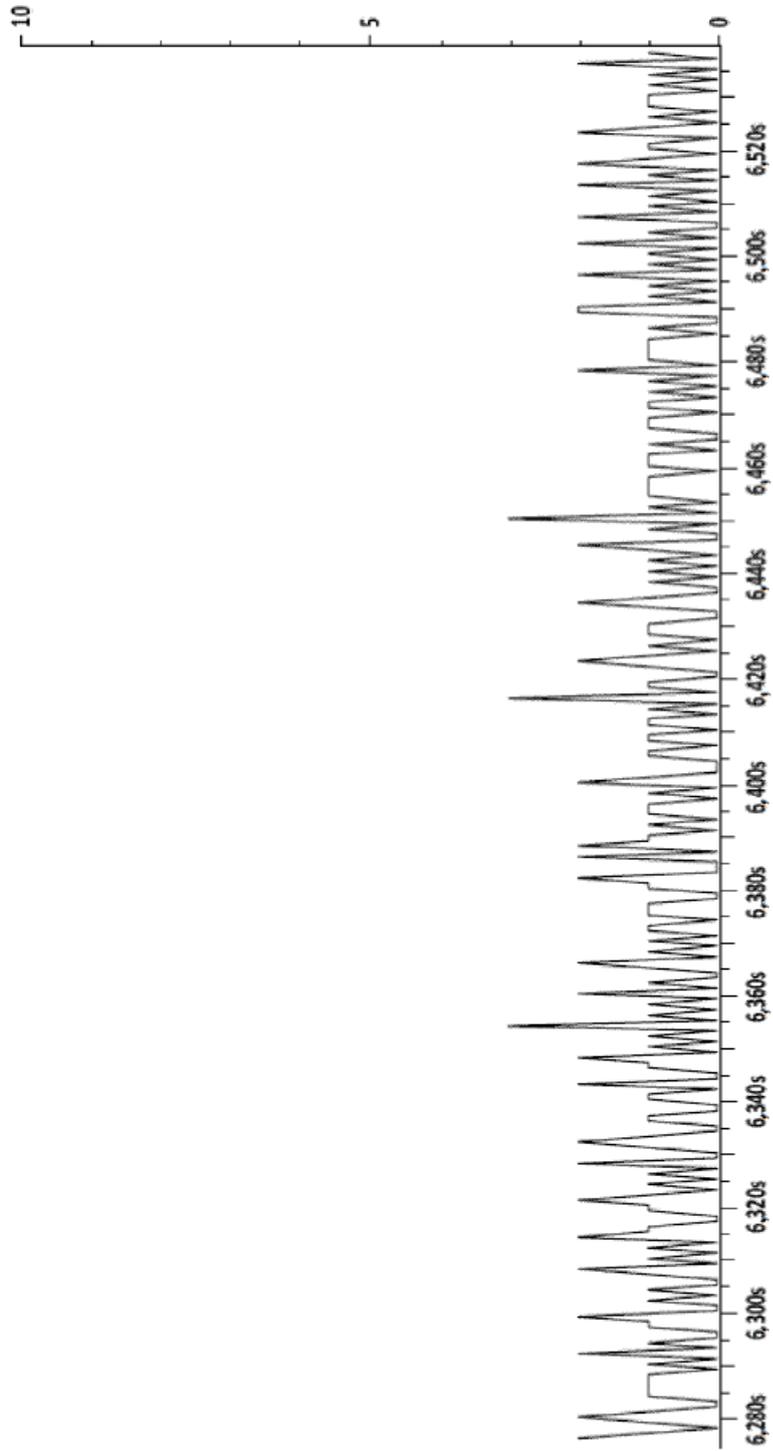


FIG. 6

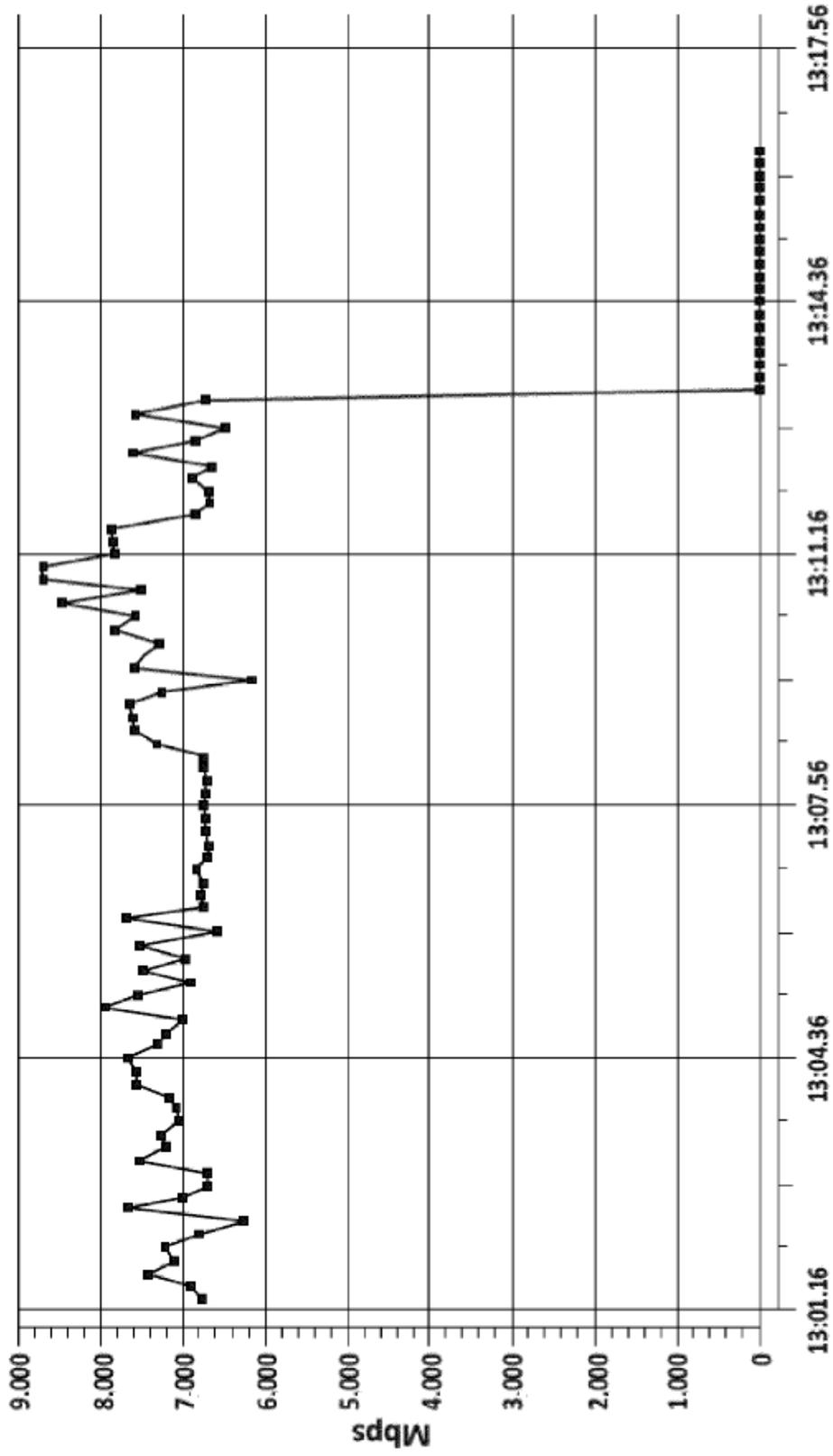


FIG. 7

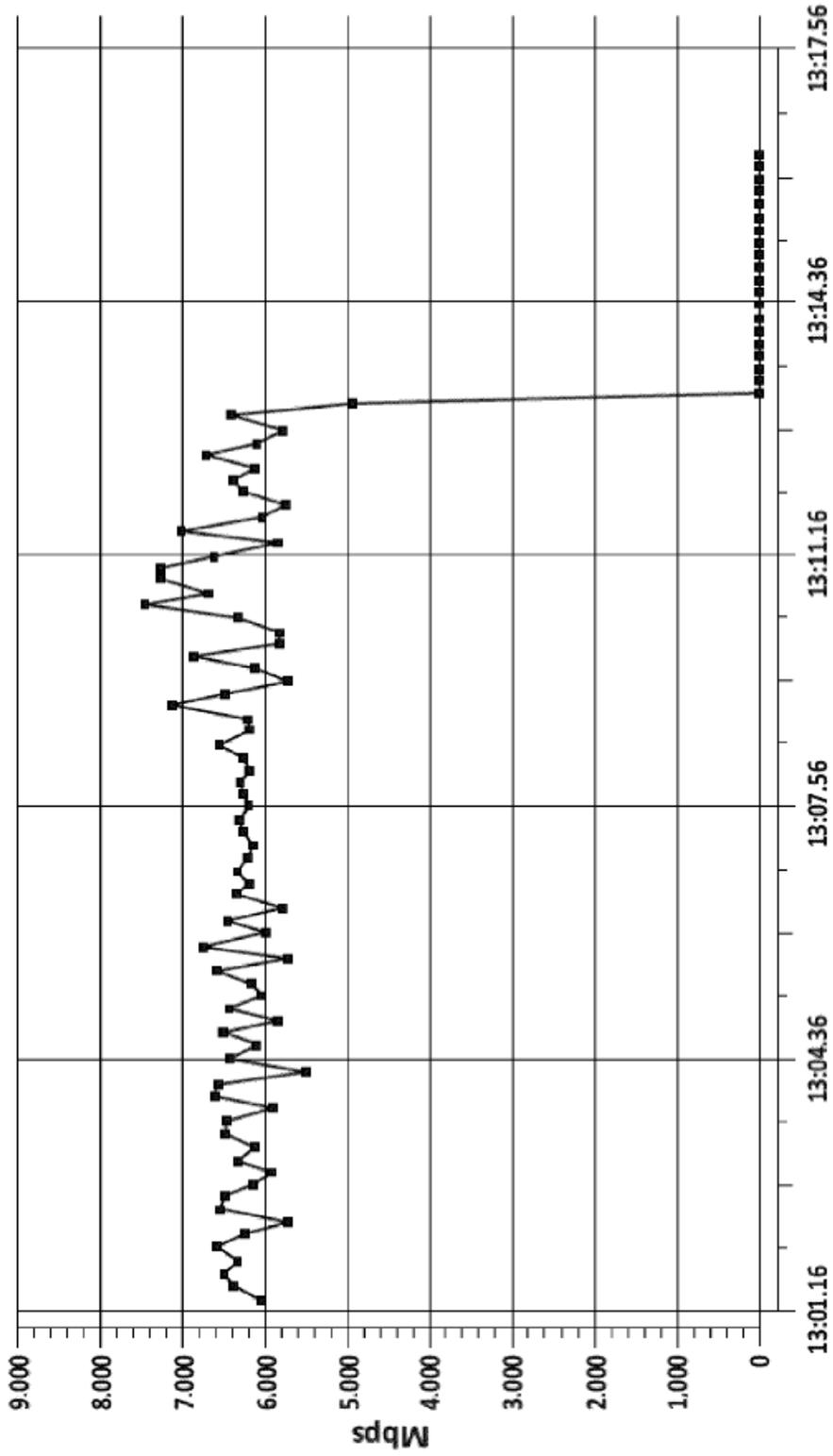


FIG. 8

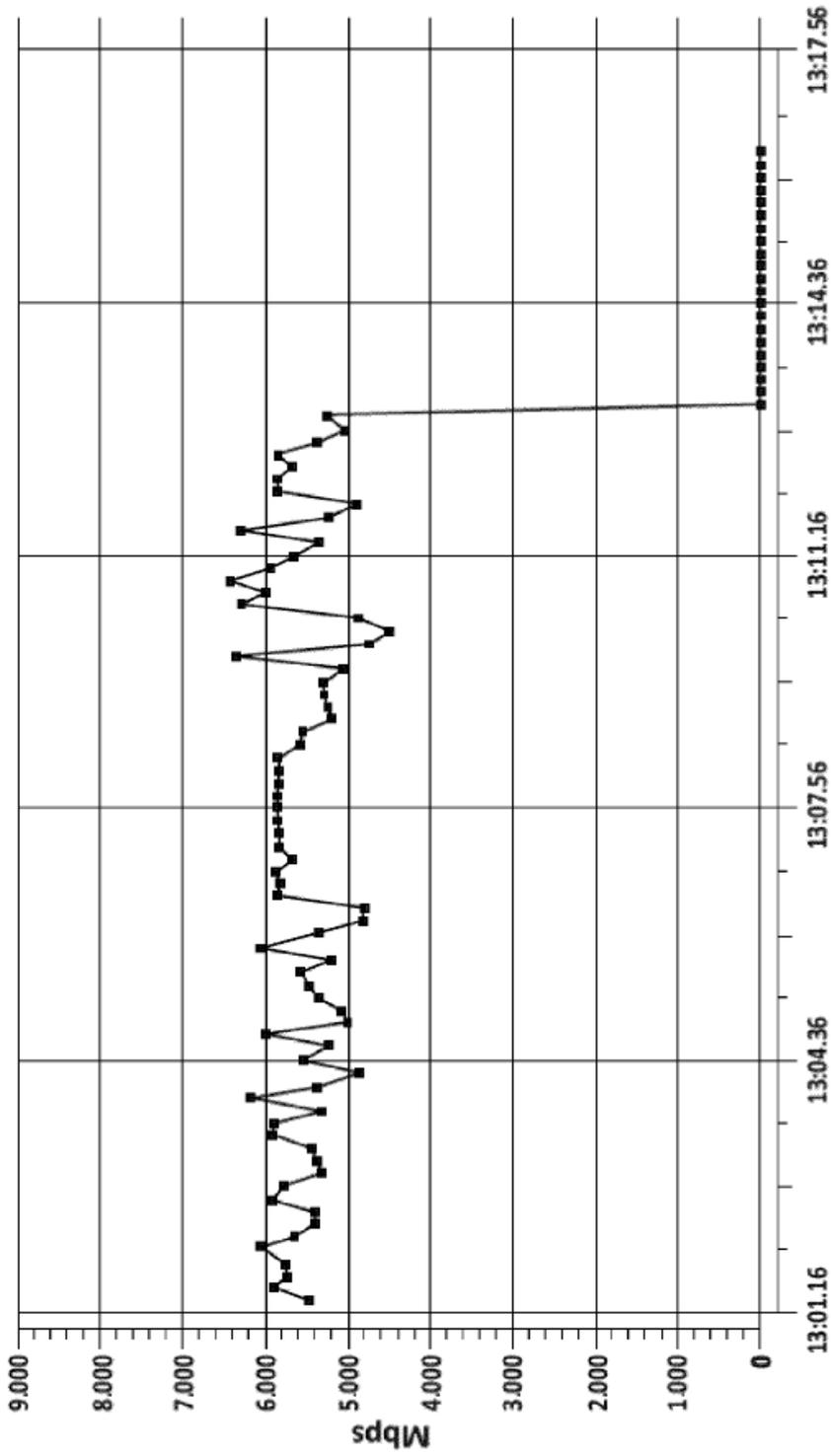


FIG. 9