# Redes de comunicaciones I Práctica 3 : Memoria

EMILIO CUESTA FERNÁNDEZ PABLO ALEJO POLANÍA BERNÁRDEZ

PAREJA 6 - GRUPO 1302

# **ÍNDICE**

1)	Introducción	P2
2)	Algunos porcentajes relevantes	,P3
3)	Direcciones IP y puertos más utilizados	P4
4)	Funciones de distribución acumulada empírica (ECDF's)	
	a) ECDF de tamaños de paquetes a nivel 2	P9
	b) ECDF de tamaños de paquetes HTTP a nivel 3	P10
	c) ECDF de tamaños de paquetes DNS a nivel 3	P11
	d) ECDF de los tiempos entre llegadas TCP	P12
	e) ECDF de los tiempos entre llegadas UDP	P13
5)	Ancho de banda a nivel 2: representación gráfica	P15
6)	Conclusiones generales	P17

## 1) Introducción:

Esta práctica consiste en el análisis detallado de la traza proporcionada por el generador de trazas dado. A lo largo de los siguientes apartados se tratará de caracterizar dicha traza hallando y representando ciertos datos de especial importancia. Estos serán de muy diversa índole, pues oscilan desde porcentajes de presencia de algunos protocolos (IP, TCP...) a gráficas que representan el ancho de banda en un determinado flujo, pasando por rankings de las direcciones y puertos más importantes y la generación de ECDFs variadas para comprender cómo se distribuyen los paquetes respecto a un determinado parámetro (Tamaño, tiempo desde el último paquete...).

Para todo ello, se han decidido emplear las siguientes herramientas:

- \*Shell Scripting: Principalmente mediante el uso de Bash
- \*Awk: De especial importancia para leer y generar datos en forma de tablas.
- \*C: Se utilizará el programa que se proporciona (CrearCDF), para generar algunos ficheros de datos.
- \*GnuPlot: Todas las representaciones gráficas presentes en este documento han sido realizadas mediante esta herramienta.

Por último, a partir de la información extraída de cada análisis puntual, se tratará de llegar a una serie de conclusiones generales, que resuman parte de la caracterización de la traza propuesta.

**Nota:** Para ejecutar el script que genera los datos sobre los que se basa este documento, basta con introducir el comando "bash practica.bash" en la terminal de su máquina.

# 2) Algunos porcentajes relevantes:

Tras ejecutar el script, al principio de la salida de terminal se muestran por pantalla los siguientes porcentajes:

```
##### APARTADO 1: PORCENTAJE DE PAQUETES IP #####

*El porcentaje de paquetes IP es 99.3352 %

De los cuales:

*El 59.7558 % son TCP

*El 6.0121 % son UDP

*El 34.2321 % son OTROS

*El porcentaje de paquetes NO-IP es 0.6648 %
```

Las cifras obtenidas concuerdan bastante con lo esperado. El protocolo IP es muy dominante, por lo que no es extraño que un 99,3352% de los paquetes estén encapsulados en el. Cabe destacar que se ha considerado como paquete IP cualquier paquete con la estructura ETH | IP ó con la estructura ETH | VLAN | IP.

El porcentaje de paquetes que no están incluidos en este conjuntos es, lógicamente, el complementario del anterior. Es fácil verificar que los dos porcentajes suman el 100%, lo que implica que no hay paquetes que no sean ni IP ni NO-IP.

Por otro lado, dentro de los paquetes IP distinguimos entre dos tipos mayoritarios. Aquellos que tienen TCP como protocolo de nivel de transporte (Casi un 60%) y aquellos que utilizan UDP (En torno a un 6%). Estos protocolos son los dos más utilizados en este nivel, por lo que no se considerará individualmente ninguno más. Si bien es cierto, los paquetes que no son ni TCP ni UDP (OTROS) tienen una fuerte presencia, sobre un 34%. En cualquier caso, TCP tiene mucha más presencia que UDP, por lo que en la traza dada se apuesta por la seguridad y garantía que ofrece este primer protocolo frente al segundo. Esto indica que en la traza no hay una mayoría de paquetes propios de un servicio de Streaming, por ejemplo la reproducción de un video en directo o de un videojuego online. En estos casos, no tiene sentido molestarse en recuperar los paquetes perdidos, por ello se utiliza mucho UDP.

Para llevar a cabo esta tarea se ha usado awk. Para filtrar los paquetes con la dirección IP esperada escribimos la sentencia "if (\$2 == 2048 || \$3 == 2048 )" siendo el campo \$2 el eth.type de la traza pcap y el campo \$3 el vlan.type. Este filtro es equivalente a lo que sería en Tshark "-Y 'eth.type eq 2048 || vlan.type eq 2048". Por otro lado para filtrar por TCP o UDP escribimos "if(\$14 == 6)" y "if(\$14 == 17)" respectivamente, siendo \$14 ip.proto. Esta escritura es equivalente en tshark a "-Y 'ip.proto eq 6"" y "-Y 'ip.proto eq 17".

# 3) Direcciones IP y puertos más utilizados:

Para representar qué direcciones y puertos son los que más se utilizan en la traza, se ha empleado el formato de "Top 10", que muestra los 10 primeros valores de acuerdo a un orden determinado. A continuación se van a mostrar los listados que se solicitan en el enunciado, es importante prestar atención a los campos **orden** y **formato** que aparecen en cada captura, pues indican cuál es el criterio para ordenar y qué magnitud tiene cada campo mostrado respectivamente.

Como introducción, utilizaremos los rankings de las direcciones IP.

```
##### APARTADO 2: TOPs 10 #####
        *TOP 10 Direcciones IP Source (Orden: Numero de paquetes)
Formato: Direccion IP//Numero de Paquetes//Numero de bytes
93.0.93.83
                 33036
                         3232149
101.196.150.194 15454
                         23098523
46.191.254.156
                         1025537
                11463
111.68.162.197
                4657
                         6918040
92.139.90.231
                 2906
                         4344112
32.141.18.135
                 2188
                         3245100
65.15.110.65
                 2161
                         3193577
68.99.78.50
                 2048
                         3009353
71.166.7.216
                 1883
                         2730262
120.106.38.240
                 1652
                         2473818
```

De esta manera, siguiendo las indicaciones del campo formato se ve que la primera columna corresponde a la dirección IP, la segunda al número de paquetes que tienen por dirección IP origen (en este caso) el valor que tienen a su izquierda en su misma fila; y la tercera al número de bytes correspondientes a esos mismos paquetes. Además, se puede apreciar que las direcciones están ordenadas decrecientemente en función del valor de la segunda columna (nº de paquetes) , como indica el campo orden. De igual manera, la siguiente imagen muestra el top 10 pero ordenado prioritariamente por los datos de la tercera columna.

```
*TOP 10 Direcciones IP Source (Orden: Numero de bytes)
Formato: Direccion IP//Numero de Paquetes//Numero de bytes
101.196.150.194 15454
                         23098523
111.68.162.197
                4657
                         6918040
92.139.90.231
                2906
                         4344112
32.141.18.135
                2188
                         3245100
93.0.93.83
                33036
                         3232149
65.15.110.65
                2161
                         3193577
68.99.78.50
                2048
                         3009353
71.166.7.216
                1883
                         2730262
120.106.38.240
                1652
                         2473818
46.191.254.156
               11463
                         1025537
```

Poco se puede decir sobre un ranking de este tipo más allá de los datos que este mismo presenta, pero es relevante ver que al ordenar bajo un criterio u otro el resultado no tiene por qué ser el mismo. Por ejemplo, puede haber direcciones que envíen muchos paquetes pero que todos ellos sean de un tamaño relativamente pequeño, de igual manera, habrá direcciones que no envíen tantos paquetes pero que trabajen con tamaños mayores. Es más, como consecuencia directa de esto, ni siquiera tienen porqué coincidir las direcciones del top10. Esto se puede apreciar en el siguiente caso, con la dirección 7.151.29.68, que aparece en el segundo ranking pero no en el primero. Por eso, las dos ordenaciones son independientes, aunque en muchos casos presenten grandes similitudes, como es lógico.

Estos son el resto de rankings generados para direcciones IP.

```
*TOP 10 Direcciones IP Dest (Orden: Numero de paquetes)
Formato: Direccion IP//Numero de Paquetes//Numero de bytes
                34986
46.191.254.156
                         50345203
101.196.150.194 3881
                         249160
124.14.247.53
                3785
                         1816645
93.0.93.83
                3076
                         2851775
111.68.162.197
                1273
                         79229
71.166.7.216
                1046
                         70017
92.139.90.231
                983
                         59576
65.15.110.65
                666
                         47886
94.141.119.155
                664
                         115206
120.106.38.240
                619
                         37073
```

```
*TOP 10 Direcciones IP Dest (Orden: Numero de bytes)
Formato: Direccion IP//Numero de Paquetes//Numero de bytes
46.191.254.156
                 34986
                         50345203
93.0.93.83
                 3076
                         2851775
124.14.247.53
                 3785
                         1816645
101.196.150.194
                 3881
                         249160
94.141.119.155
                 664
                         115206
111.68.162.197
                         79229
                 1273
7.151.29.68
                 227
                         76301
71.166.7.216
                 1046
                         70017
92.139.90.231
                 983
                         59576
65.15.110.65
                 666
                         47886
```

A continuación se presentan los rankings correspondientes a los **puertos TCP**:

```
*TOP 10 Puerto TCP Source (Orden: Numero de paquetes)
Formato: Puerto TCP//Numero de Paquetes//Numero de bytes
        36640
80
                52857665
55934
        1423
                88065
55860
        1096
                67367
54615
        1046
                70017
55865
        617
                40574
43585
        607
                36512
        603
33896
                35533
55173
        471
                28338
55848
        418
                26181
                24043
33903
        380
```

```
*TOP 10 Puertos TCP Source (Orden: Numero de bytes)
Formato: Puerto TCP//Numero de Paquetes//Numero de bytes
80
        36640
                52857665
443
        229
                217800
55934
        1423
                88065
54615
        1046
                70017
55860
        1096
                67367
        617
55865
                40574
43585
        607
                36512
33896
        603
                35533
55173
        471
                28338
46832
        52
                26382
```

```
*TOP 10 Puerto TCP Dest (Orden: Numero de paquetes)
Formato: Puerto TCP//Numero de Paquetes//Numero de bytes
80
        12356
                1074339
55934
        5486
                8236507
55860
        4313
                6437994
55865
        3204
                4808618
                3245100
43585
        2188
54615
        1883
                2730262
33896
        1813
                2707440
        1717
55173
                2566453
55848
        1396
                2072650
       1174
                1756652
46371
```

```
*TOP 10 Puertos TCP Dest (Orden: Numero de bytes)
Formato: Puerto TCP//Numero de Paquetes//Numero de bytes
55934
        5486
                8236507
55860
        4313
                6437994
55865
        3204
                4808618
43585
        2188
                3245100
54615
        1883
                2730262
33896
        1813
                2707440
55173
        1717
                2566453
                2072650
55848
        1396
46371
        1174
                1756652
57063
        1133
                1690967
```

Por último, se exponen los rankings de puertos UDP:

```
*TOP 10 Puerto UDP Source (Orden: Numero de paquetes)
Formato: Puerto UDP//Numero de Paquetes//Numero de bytes
48883
        3785
                 1816645
53
        592
                 85720
546
        124
                 18337
5353
        95
                 23317
1900
        12
                 6447
63423
        6
                 1080
58532
        6
                 1080
55421
        6
                 1080
49169
        6
                 1080
61153
        3
                 624
```

```
*TOP 10 Puertos UDP Source (Orden: Numero de bytes)
Formato: Puerto UDP//Numero de Paquetes//Numero de bytes
48883
        3785
                1816645
53
        592
                85720
5353
        95
                23317
546
        124
                18337
1900
        12
                6447
63423
        6
                1080
58532
        6
                1080
55421
        6
                1080
49169
        6
                1080
61153
        3
                624
```

```
*TOP 10 Puerto UDP Dest (Orden: Numero de paquetes)
Formato: Puerto UDP//Numero de Paquetes//Numero de bytes
4939
53
        3785
                 1816645
        591
                 46391
5355
        134
                 11460
547
        124
                 18337
5353
        95
                 23317
1900
        42
                 12015
8000
        2
                 190
        2
5035
                 461
9920
        1
                 92
9800
        1
                 108
```

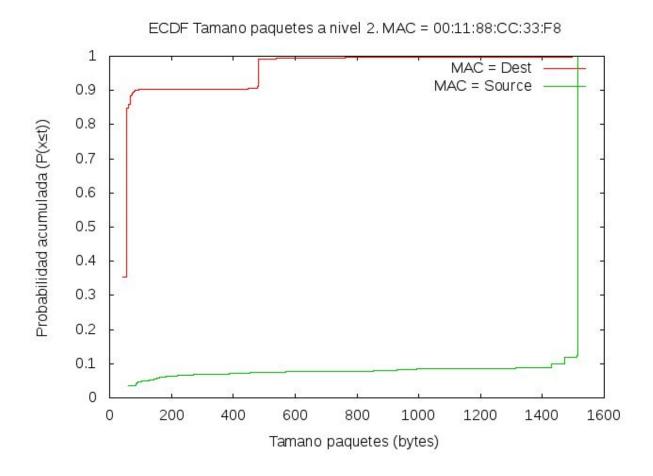
```
*TOP 10 Puerto UDP Dest (Orden: Numero de bytes)
Formato: Puerto UDP//Numero de Paquetes//Numero de bytes
4939
        3785
                 1816645
53
        591
                46391
5353
        95
                 23317
547
        124
                18337
1900
        42
                12015
        134
5355
                11460
5035
        2
                461
64925
        1
                 394
        1
23710
                 318
34968
        1
                 316
```

### 4) Funciones de distribución acumulada empírica (ECDF's)

En los proximos subapartados se muestran las ECDF's de distintas características de los paquetes (por ejemplo, el tamaño o el tiempo entre llegadas) mediante gráficas. Para ello se trabaja sobre ciertos paquetes, resultado de un filtro sobre sus direcciones MAC, HTTP o DNS, o mediante sus puertos TCP y UDP. Estas gráficas reflejan la probabilidad de que un paquete filtrado tenga un tamaño (o menos) o haya llegado en un tiempo determinado (o antes), dependiendo del caso. Por tanto, suponen una herramienta de análisis del tráfico de red muy útil.

#### 4.1) ECDF de tamaños de paquetes a nivel 2

La siguiente gráfica es la asociada a la ECDF del **tamaño a nivel 2** de los paquetes de la traza con **dirección MAC = 00:11:88:CC:33:F8**:



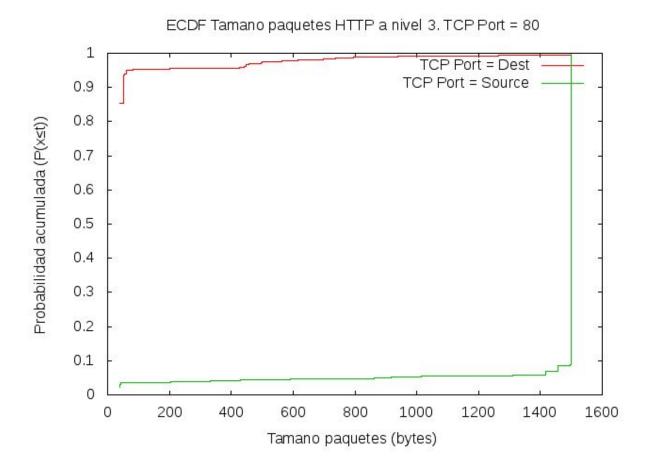
Las direcciones **MAC** sirven para identificar un dispositivo de red. Por tanto son únicas y vienen definidas desde la fabricación de cada dispositivo.

Mirando la gráfica se puede apreciar que el 90% de los paquetes recibidos por el dispositivo al que corresponde esa MAC (Paquetes MAC destino, línea roja) son menores a los 100 bytes

y por el contrario, menos del 15% de los paquetes enviados por ese dispositivo (Paquetes MAC origen, línea verde) son menores a los 1500 bytes. Esto nos indica que el dispositivo con esta MAC es un dispositivo que recibe una gran cantidad de peticiones (de tamaño reducido) y envía respuestas con un tamaño considerable (información requerida por los clientes, generalmente). Se comportaría de manera similar a esta gráfica un servidor, por ejemplo uno de Wikipedia.

#### 4.2) ECDF de tamaños de paquetes HTTP a nivel 3

La siguiente gráfica está asociada a la ECDF del **tamaño a nivel 3** de los paquetes **HTTP** (puerto TCP = 80):

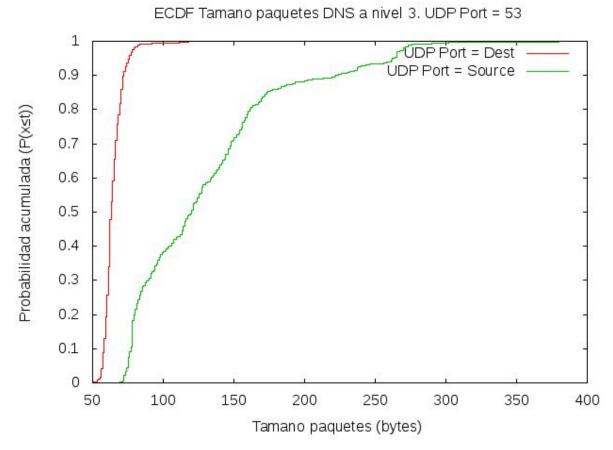


HTTP es un protocolo caracterizado por su estructura petición-respuesta entre cliente y servidor. El análisis de esta gráfica es análogo al anterior. Como se puede apreciar, en el caso de puerto TCP origen, la gráfica describe un crecimiento apenas perceptible hasta los 1500 bytes, donde crece de golpe. Esto refleja que menos del 10% de los paquetes tienen tamaño inferior a los 1500 bytes. Este hecho indica que dichos paquetes son las respuestas del servidor al cliente, debido a que la mayoría de estos son de tamaño elevado (llevan gran cantidad de información) al transportar el recurso pedido. Por el contrario, la función de los paquetes con puerto TCP destino muestra que el 95% son de tamaño reducido (menos de 100

bytes), acorde a lo esperado de una petición de recursos del cliente a un servidor. También, los paquetes de menor tamaño, se pueden ver como paquetes enviados por el cliente para confirmar la conexión con el servidor. Un ejemplo que daría lugar a esta gráfica sería el de una red social, en la que el cliente pide información sobre otros usuarios y se le envían los datos requeridos.

#### 4.3) ECDF de tamaños de paquetes DNS a nivel 3

La siguiente gráfica está asociada a la ECDF del tamaño a nivel 3 de los paquetes DNS (puerto UDP = 53):



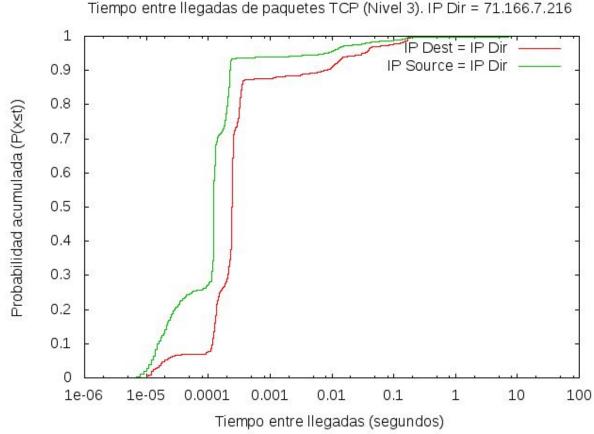
El protocolo **DNS** (Domain Name System) se encarga, en su mayoría, de "traducir" direcciones IP o la localización de servidores de correo electrónico a "nombres legibles" para el ser humano (URL, dirección de correo, etc). Un ejemplo de ello sería la "traducción" de la dirección IP 218.12.4.56 al "nombre" www.google.es. Esto se consigue mediante unas bases de datos asociadas a los nombres de dominio (domain name) de los dispositivos conectados a una red. Así, los dispositivos, son direccionados.

Es importante mencionar la diferencia que hay, aunque solo se hayan comprobado dos casos puntuales, entre el tamaño de los paquetes DNS filtrados por UDP (400 bytes como mucho) y el tamaño de los paquetes TCP de la anterior gráfica (alcanzan los 1500 bytes). Esta diferencia se debe a que los paquetes TCP proveen conexión extremo a extremo necesitando

así gran cantidad de información en cada paquete y los UDP no, haciendo así sus paquetes más "livianos". Un servidor DNS que genere esta gráfica sería el que guarde localizaciones de servidores de correo electrónico. Recibiendo así paquetes muy pequeños (dirección de correo electrónico) y enviando paquetes con mayor información con la "traducción" a su identificador binario. Por ejemplo @gmail.com.

#### 4.4) ECDF de tiempo entre llegadas TCP

La siguiente gráfica está asociada a la ECDF del tiempo entre llegadas de los paquetes con **protocolo TCP** y con **IP** = **71.166.7.216**, tanto de destino como de origen:

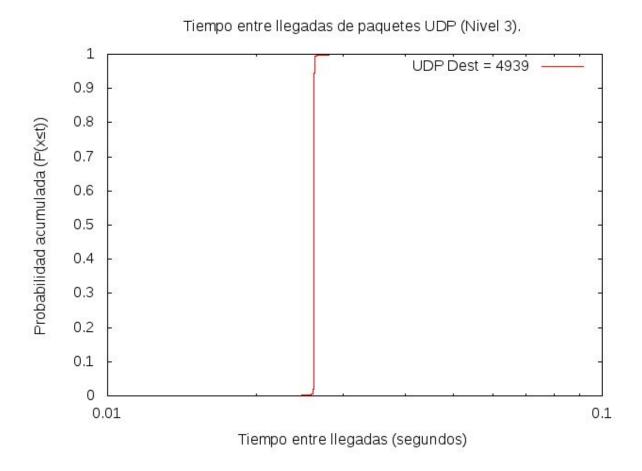


El protocolo **TCP** sirve para permitir que un equipo esté conectado a una red. Para ello provee una conexión segura de extremo a extremo entre cliente y servidor, asegurándose de que los paquetes llegan a su destino y en su correcto orden. Un ejemplo de ello sería el acceso a una red social. en la que el servidor pide un usuario y contraseña, el cliente las envía y el servidor comprueba los datos, permitiendo o no el acceso a la página. Esta sucesión de acciones se lleva a acabo siempre en ese orden y sin perder información en la transmisión de datos. Se puede apreciar que la diferencia entre las gráficas de cada función de distribución es mínima y que en ambas el 90% de los paquetes tardan menos de 1 ms desde el anterior. Esto se debe a la importancia de la conexión extremo a extremo tanto para el cliente como para el servidor y de la necesidad de una transmisión de la respuesta casi inmediata. El tiempo con IP

destino puede ser mayor debido a cierta congestión. En el caso de que la dirección IP esté asociada a un servidor, éste puede tener varias peticiones de distintos clientes lo que implica un retraso en su recepción (congestión). Una vez asimilada la petición no tardará tanto tiempo en enviar la respuesta. Un servicio que pudiera generar esta gráfica sería la transmisión de mensajes de texto por WhatsApp.

#### 4.5) ECDF de tiempo entre llegadas UDP

La siguiente gráfica está asociada a la ECDF del **tiempo entre llegadas** de los paquetes con **protocolo UDP = 4939**:

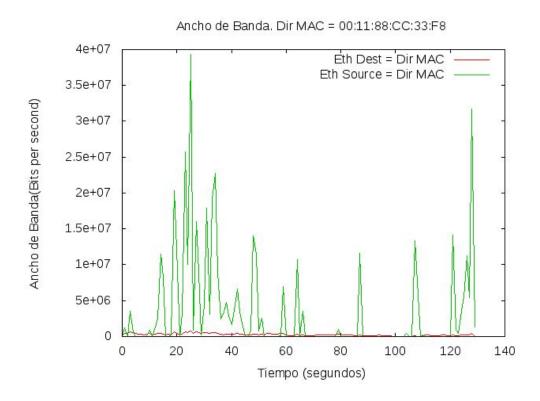


El protocolo **UDP** permite enviar paquetes a través de la red sin haber establecido una conexión previa con el destino. Los paquetes encapsulados en este protocolo tampoco se reenvía en caso de pérdidas. Esto hace que UDP no sea muy fiable, pudiendo perder información en el trayecto. Mirando la gráfica vemos que solo hay función de distribución en el puerto UDP destino y que además los tiempos entre llegadas se centran en un solo valor. Esto indica por un lado, que ese puerto no envía paquetes UDP y por el otro, que la transmisión de paquetes de este tipo es "constante", en el sentido de que todos los paquetes llegan con la misma diferencia de tiempo respecto a la llegada del anterior. Además el tiempo entre paquetes no supera el 0,1 s, lo que permitiría que la información transmitida

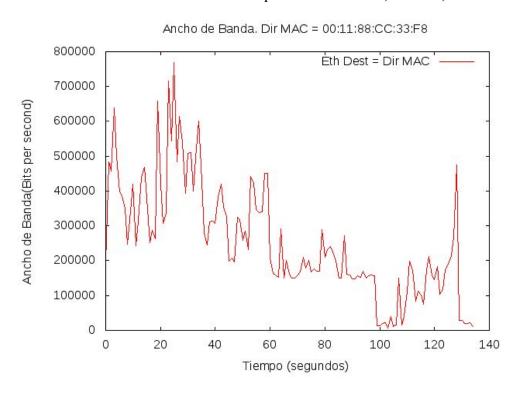
perteneciera a una transmisión en vivo. Este es uno de los servicios que más utilizan UDP. Por ello, algo que podría generar esta gráfica sería la visualización de un video en vivo o streaming (Twitch). En ella el cliente con ese puerto destino UDP recibe paquetes pero no necesita confirmar al dispositivo que se los ha enviado que los está recibiendo. De ahí la ausencia de paquetes con ese puerto UDP origen. Que el tiempo entre llegadas sea el mismo (además de pequeño) es importante ya que en una transmisión en vivo se quiere evitar cualquier tipo de jitter que disminuya la calidad del servicio. En relación con una retransmisión de un partido de fútbol, no es agradable estar viendo el partido a ritmo normal y que de repente se "congele" la imagen, teniendo que esperar un tiempo, por pequeño que sea.

# 5) Ancho de banda a nivel 2: representación gráfica

La siguiente gráfica está asociada al **ancho de banda (bps)** de los paquetes con **dirección MAC** = **00:11:88:CC:33:F8**:



Para una mejor visualización de los datos, también se adjunta una imagen ampliada del comportamiento de los ancho de banda de recepción de esa MAC, es decir, de la línea roja.



Sabiendo que la dirección MAC es única para cada dispositivo, vemos que nuestro dispositivo emite una gran cantidad de datos (MAC origen) y recibe una pequeña cantidad (MAC destino). Esto resulta evidente viendo la diferencia de magnitudes en la primera gráfica. Se comportaría así un servidor recibiendo peticiones (las cuales no requieren de mucha información en sus paquetes) y enviando los datos requeridos (mayor número de bytes). Normalmente a cada petición le corresponde una respuesta de tamaño mayor, eso explica por qué la línea verde está siempre por encima de la roja. Es interesante en estas gráficas fijarse en los picos, ya que permiten saber cuando hay más tráfico en la red y poder predecir posibles congestiones, necesidad de recursos, etc... Un ejemplo válido sería el de un periódico online como Marca.com, que seguramente sea más consultado a unas horas del día que a otras. Es decir, tendrá un tráfico total muy variable, desde momentos de alta demanda a momentos de muy poca actividad (véase la segunda mitad del intervalo 60-80 en la primera gráfica de este apartado).

### 6) Conclusiones generales

Esta práctica muestra la utilidad de herramientas como awk y GNUplot. En el caso de awk, utilizado en conjunto con ciertos comandos de shell scripting, es evidente su efectividad como herramienta de análisis de ficheros. De igual manera, GNUplot resulta un instrumento muy útil para la representación, facilitando la posterior visualización, de los datos obtenidos de la traza.

En líneas generales, la práctica ha servido para incrementar en gran medida los conocimientos relativos al análisis de tráfico de red. Por ejemplo, la diferencia de tamaño de paquetes entre protocolos de transporte con conexión extremo a extremo (TCP) y sin ella (UDP) ha quedado perfectamente retratada. Los numero de bytes enviados en el primer caso, por la necesidad de confirmar que el paquete ha llegado con éxito, es muy superior al del segundo.

Cada una de las gráficas aporta información distinta. Por ejemplo, la gráfica 4.5 es de especial interés: muestra la necesidad de un tiempo de transmisión prácticamente constante en algunas aplicaciones online que se sirven del protocolo UDP, como los streamings. Por otro lado, la gráfica 4.2 deja claro la diferencia de tamaño entre los paquetes de una petición (rondando los 100 bytes) y los de una respuesta (1500 bytes) en el caso de HTTP.

Por tanto, gracias a las gráficas es posible diferenciar la naturaleza de la conexión entre cliente y servidor a través de distintos filtros. Es más, analizando el tráfico de red propio de los paquetes encapsulados en un protocolo u otro, vemos cómo hay protocolos que determinan el comportamiento de todo un flujo, como UDP. También hay otros más generales como IP o ETH que dada su alta presencia (Consultar porcentajes del apartado 1) presentan una información menos concreta.

Además, el uso de ECDFs para representar los distintos datos facilita en gran medida el hecho de analizar la distribución de los paquetes que pasan un determinado filtro.

En cuando al ancho de banda, último apartado de la entrega, se ha comprobado que la dirección MAC empleada envía una cantidad de información muy superior a la que recibe en casi todo momento. Esto lleva a pensar que esta dirección MAC pertenece a un servidor, que envía mucha información pero recibe muy poca, principalmente peticiones.

Como conclusión y en relación con lo expuesto anteriormente, en todas las representaciones gráficas se observa muy bien cómo en muchos casos un gran % de los paquetes (70, 80...%) de un mismo tipo tienen propiedades bastante comunes. Es decir, paquetes propios de servicios similares tienen, en general, características similares.