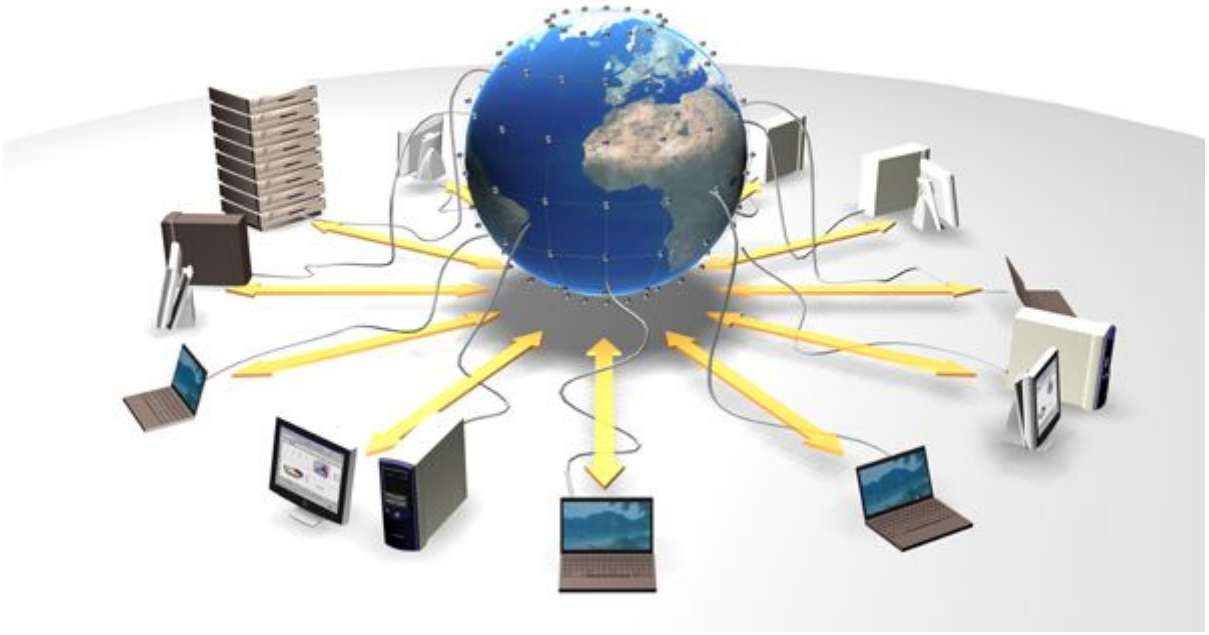
MEMORIA

Práctica 3: Monitorización



Andrés Salas Peña - andes.salas@estudiante.uam.es
Francisco Vicente Lana - francisco.vicentel@estudiante.uam.es

1. Introducción

Tras haber desarrollado las dos primeras prácticas cuyo objetivo era mostrar al alumno las herramientas básicas de las que se dispone (Wireshark, Libpcap) para analizar los distintos paquetes que maneja una red, en esta tercera práctica de la asignatura de Redes de Comunicaciones I se propone ir un paso más allá de forma que consiste en una introducción en la monitorización de redes con el fin de analizar en profundidad el rendimiento y estado para poder implementar mejoras de cara al futuro. Para este fin se nos propone la creación de un script con el que obtener diversos indicadores del funcionamiento de una red mediante el uso de la herramientas como tshark, awk o GNUplot.

Nuestro enfoque ha consistido en hacer diversas pruebas preliminares para dominar ciertos comandos útiles de tshark para obtener los datos deseados sobre la traza obtenida mediante el generador proporcionado, así como comprender la funcionalidad ofrecida por awk para intuir cómo ambas herramientas combinadas nos permiten realizar los análisis deseados.
Una vez realizada esta primera fase, nos dispusimos a crear un único script (analizador.sh) que generase los ficheros .txt necesarios para analizar el tráfico de la traza proporcionada. Estos ficheros son generados por **tshark**, mientras que **awk** los manipula con el fin de imprimir por pantalla los resultados perseguidos o generar un gráfico descriptivo de cada caso mediante GNUplot.
Una complicación a destacar ha sido el cálculo de las funciones de distribución acumulada empírica (ECDFs); una vez habíamos comprendido su definición en vez de utilizar el fichero auxiliar crearCDF.c, decidimos implementar el cálculo de estas directamente en el script manipulando los datos en ficheros .txt por medio de akw para finalmente obtener la muestra a representar por GNUplot.

Por último, cabe mencionar que el código de nuestro script ha sido estructurado en cinco puntos modulares los cuales se encuentran juntos de cara a la entrega de un único ".sh". Son respectivamente: Porcentaje de protocolos, Tops de direcciones y puertos, ECDFs de tamaños, ECDFs de tiempos (interarrivals) y Serie temporal.

La siguiente imagen muestra los datos de la traza correspondiente a nuestra pareja:

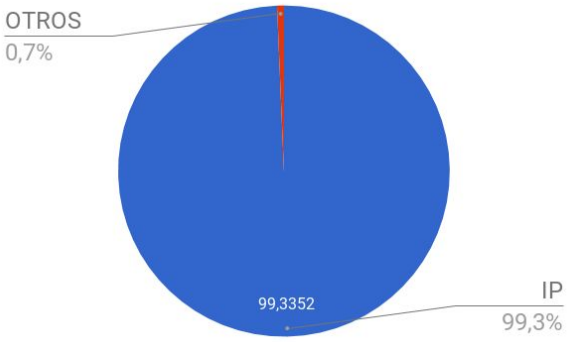
```
Generando traza (trazando.pcap) para la pareja 7 del grupo 1302. (id=130207)
Epoch de inicio: 1510696095

Información útil para el estudiante:
La dirección MAC que deberá tener en cuenta es: 00:11:88:CC:33:FC
La dirección IP (origen o destino) del flujo TCP que deberá tener en cuenta es:
53.59.245.211
El puerto (origen o destino) del flujo UDP que deberá tener en cuenta es: 1944

PCAP generado con éxito!
```

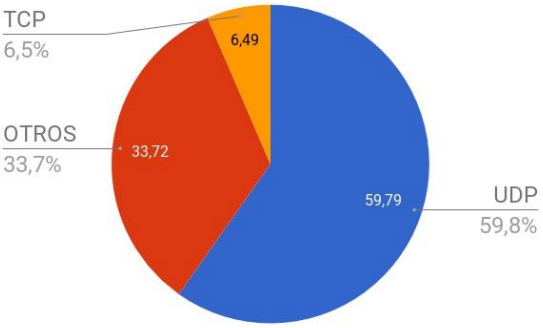
2. Análisis de Resultados

2.1 Porcentaje de Paquetes

En este primer apartado se nos pide dar porcentajes generales sobre los protocolos de los paquetes incluidos en nuestra traza.

```
PUNTO 1 (porcentajes protocolos) **********************************
Porcentaje de paquetes IP: 99.3352% (No IP: 0.664775%)
Entre los IP se tiene:
        UDP: 6.49256%
        TCP: 59.7897%
        Otros: 33.7178%
```





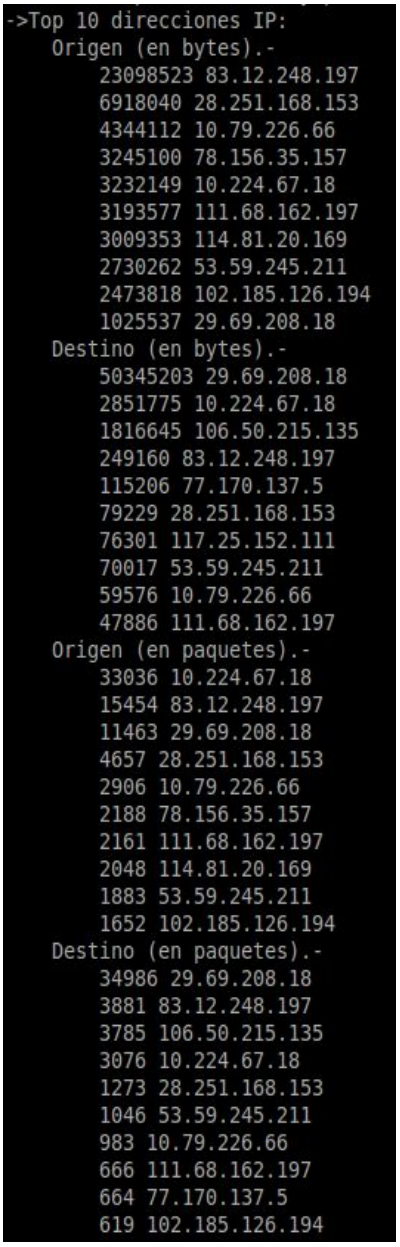
Observando los datos obtenidos podemos concluir que la amplía mayoría de paquetes capturados presentan el protocolo IP (o bien un VLAN seguido de IP) en su tercer nivel. Dentro de los paquetes con IP, se aprecia que más de la mitad de los paquetes son TCP mientras que una pequeña cantidad de ellos son UDP a cuarto nivel. Los filtros utilizados para cada caso son los siguientes:

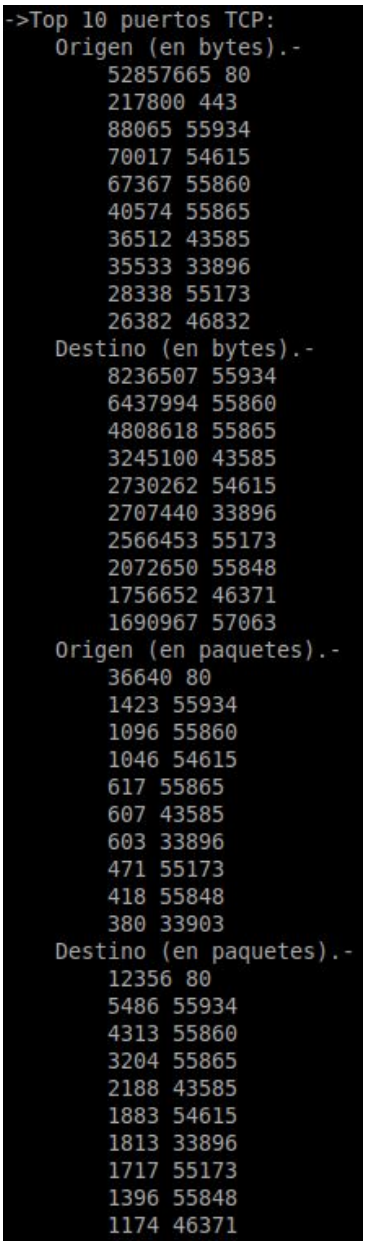
IP: 'ip'
UDP: ‘udp’
TCP: ‘tcp’

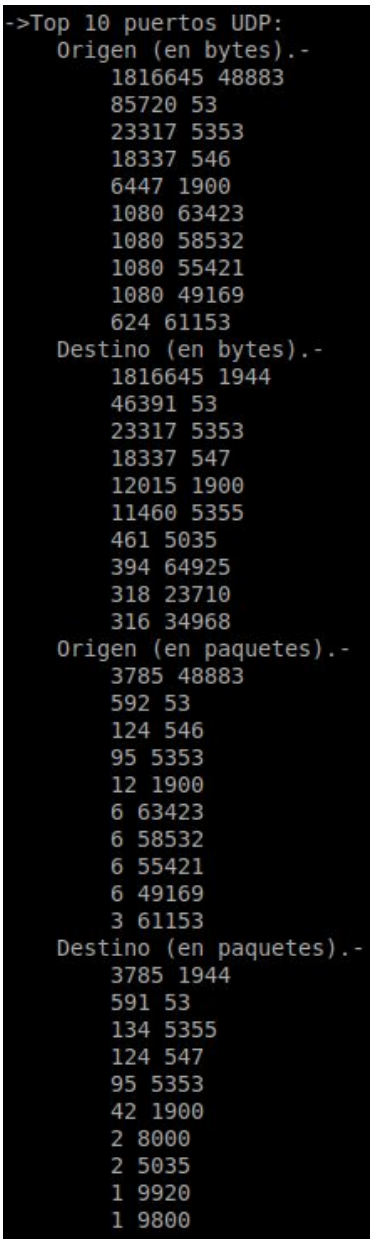
Mediante tshark utilizamos estos filtros con el fin de crear ficheros con las direcciones/puertos y la longitud del paquete (útil para los tops por bytes) como columnas para después por medio de awk y el comando word count obtener las estadísticas deseadas.

2.2 Top 10 IP Activas y Puertos Activos

A continuación adjuntamos capturas de pantalla de la salida de nuestro script que muestran los diferentes top's exigidos:





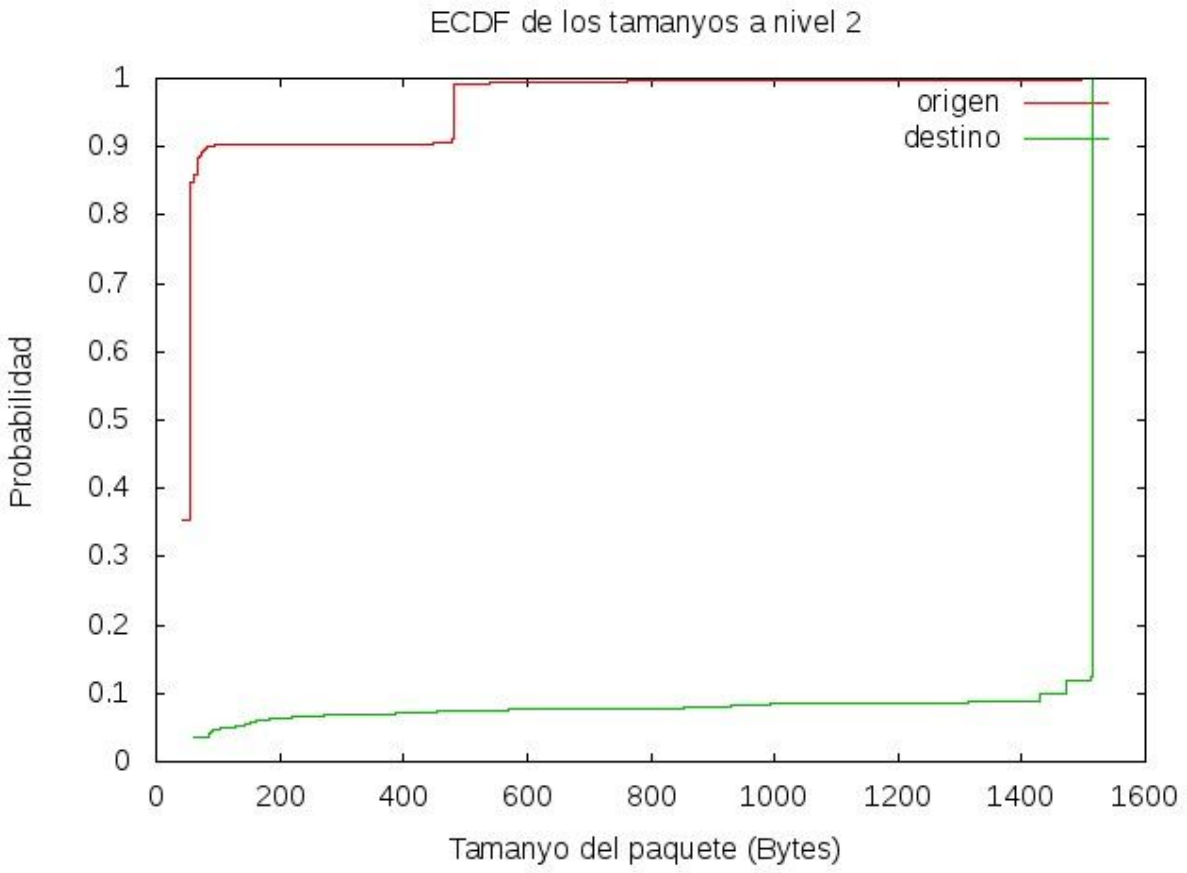


En esta apartado no hay mucho que analizar. Simplemente habría que aclarar que la primera columna de cada listado se corresponde con el número de paquetes o bytes de cada dirección o puerto que se muestran en la segunda columna.

Por lo demás, destaca la dirección IP 29.69.208.18 como la de destino más común tanto en bytes como en paquetes. Pasa lo mismo en el caso del puerto de origen 80 por parte de TCP (se distancia mucho del resto) y además también gana a todos como puerto de destino en cuanto a número de paquetes (se trata de paquetes de pequeño tamaño, ya que no aparece en el top por bytes). Respecto a UDP podemos apreciar que los puertos 48883 y 1944 son los más repetidos como origen y destino respectivamente.

2.3 ECDF Tamaños Nivel 2

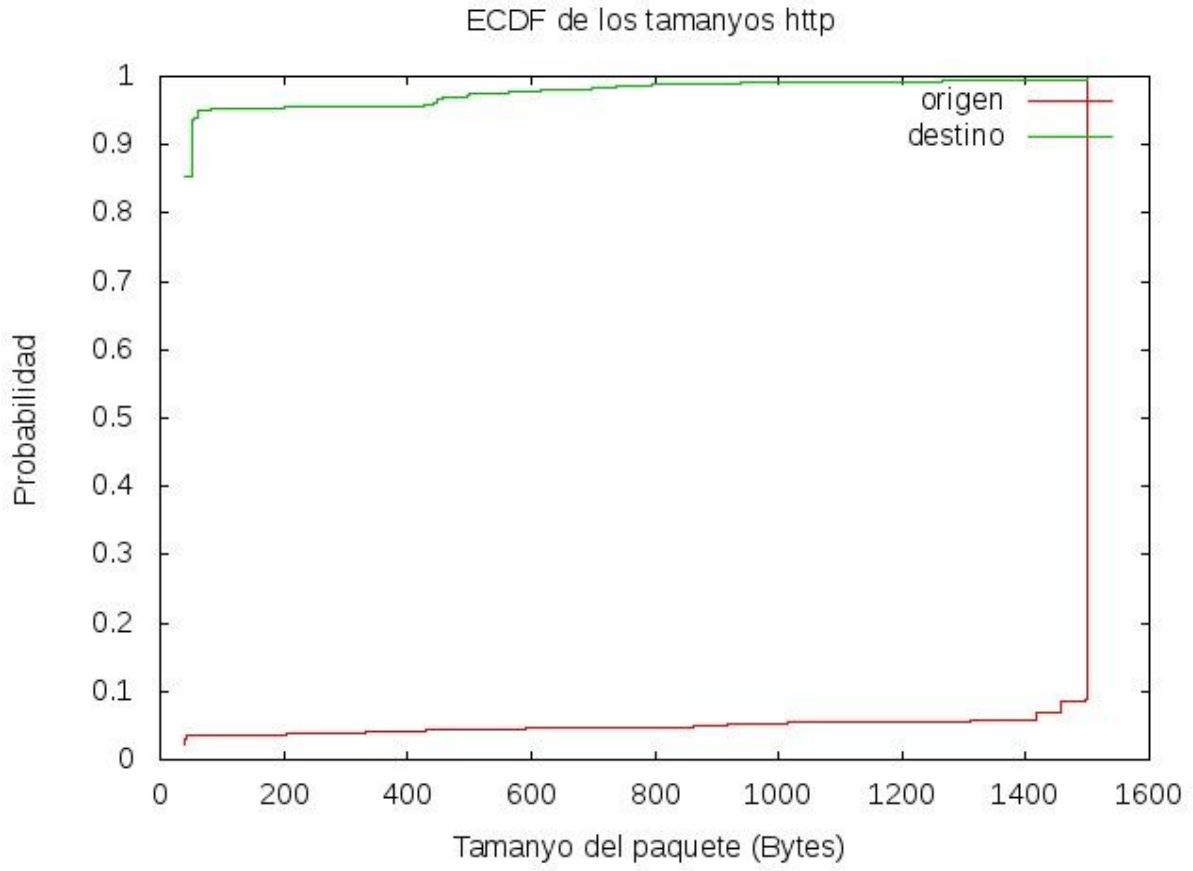
En este apartado se nos pide, utilizando la dirección MAC proporcionada por el generador (00:11:88:CC:33:FC), representar el tamaño a nivel dos de los paquetes. La gráfica que se muestra a continuación representa estos datos. La línea roja describe el flujo de salida ya que se ha filtrado para que tan solo se tengan en cuenta los paquetes cuya dirección de origen ethernet coincida con la MAC anterior; mientras que la verde se corresponde con el flujo de entrada al ser paquetes con la dirección de destino dada.



Gracias a esta ECDF podemos ver y entender al instante el comportamiento de la red a analizar. Se aprecia que los paquetes salientes son en general muy cortos y sin embargo los de entrada son en su mayoría largos (la línea vertical al final de la función verde nos permite deducir que casi todos los estos paquetes tienen la longitud máxima de 1514 bytes). Se puede intuir que esto es debido a que en general nos encontramos con que las comunicaciones cliente-servidor suelen ser pequeñas al tratarse fundamentalmente de peticiones, frente a las servidor-cliente que son más grandes ya que suelen consistir en datos. Por ello en este caso la MAC que nos ocupa debe estar asociada a un dispositivo “cliente”.

2.4 ECDF Tamaños Nivel 3 de HTTP

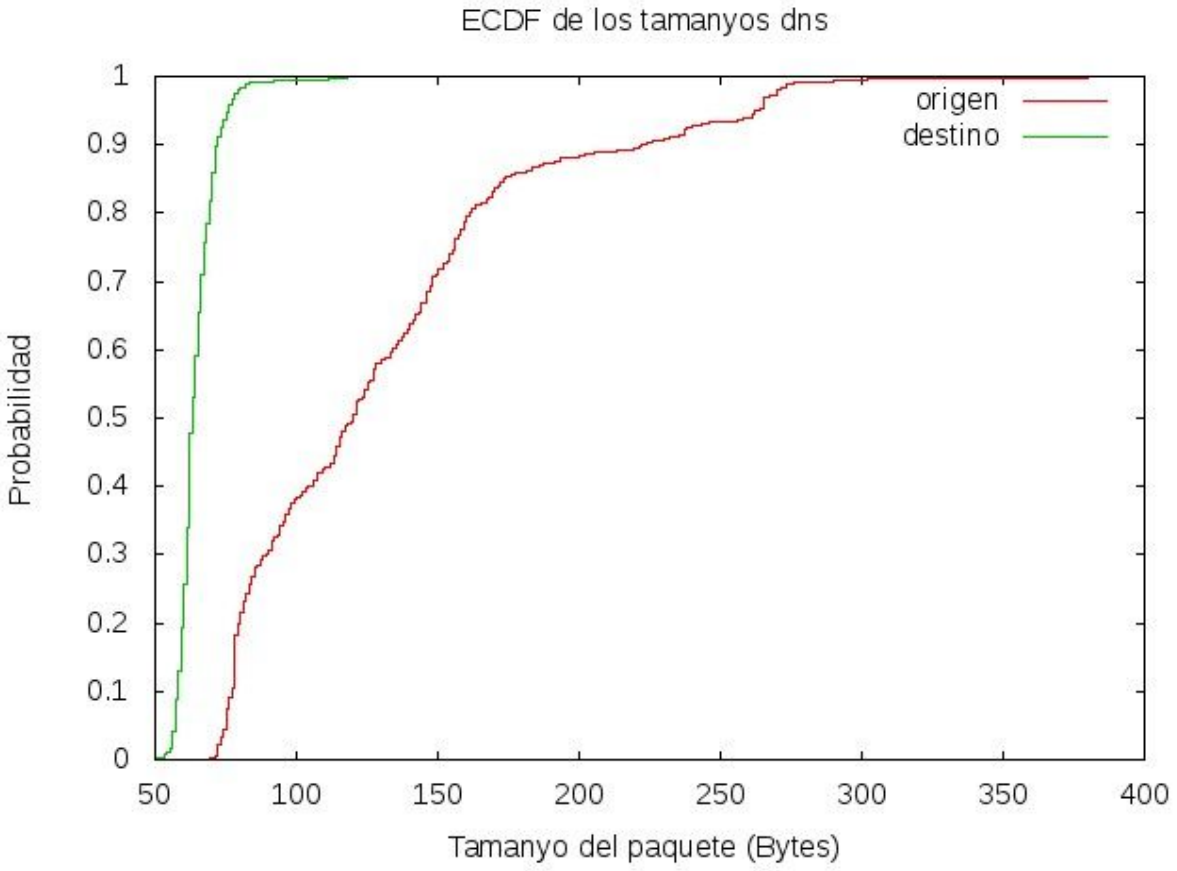
En este apartado se nos pide, utilizando que los paquetes HTTP son aquellos que utilizan el puerto 80 de TCP en origen o destino, representar el tamaño a nivel tres de los paquetes. La gráfica que se muestra a continuación representa estos datos. La línea roja describe el flujo de salida ya que se ha filtrado para que tan solo se tengan en cuenta los paquetes cuyo puerto de origen TCP coincida con el valor 80 anterior; mientras que la verde se corresponde con el flujo de entrada al ser paquetes con el puerto de destino dado.



Mediante esta ECDF nos hacemos a la idea en seguida del comportamiento de la red a analizar. Se aprecia que los paquetes salientes son en general largos y sin embargo los de entrada son en su mayoría cortos. Esto es debido a que el protocolo HTTP es un protocolo cliente-servidor que se basa en operaciones de solicitud/respuesta. Los paquetes salientes (correspondiente a la línea roja y comunicación cliente-servidor) son de gran tamaño ya que el cliente por lo general envía una URL (formada por identificador de protocolo de acceso, dirección DNS o IP del servidor, puerto y objeto requerido). Por otro lado, los paquetes entrantes (correspondientes con la línea verde y comunicación servidor-cliente) son bastante pequeños porque el servidor solo envía una respuesta con un código de estado.

2.5 ECDF Tamaños Nivel 3 de DNS

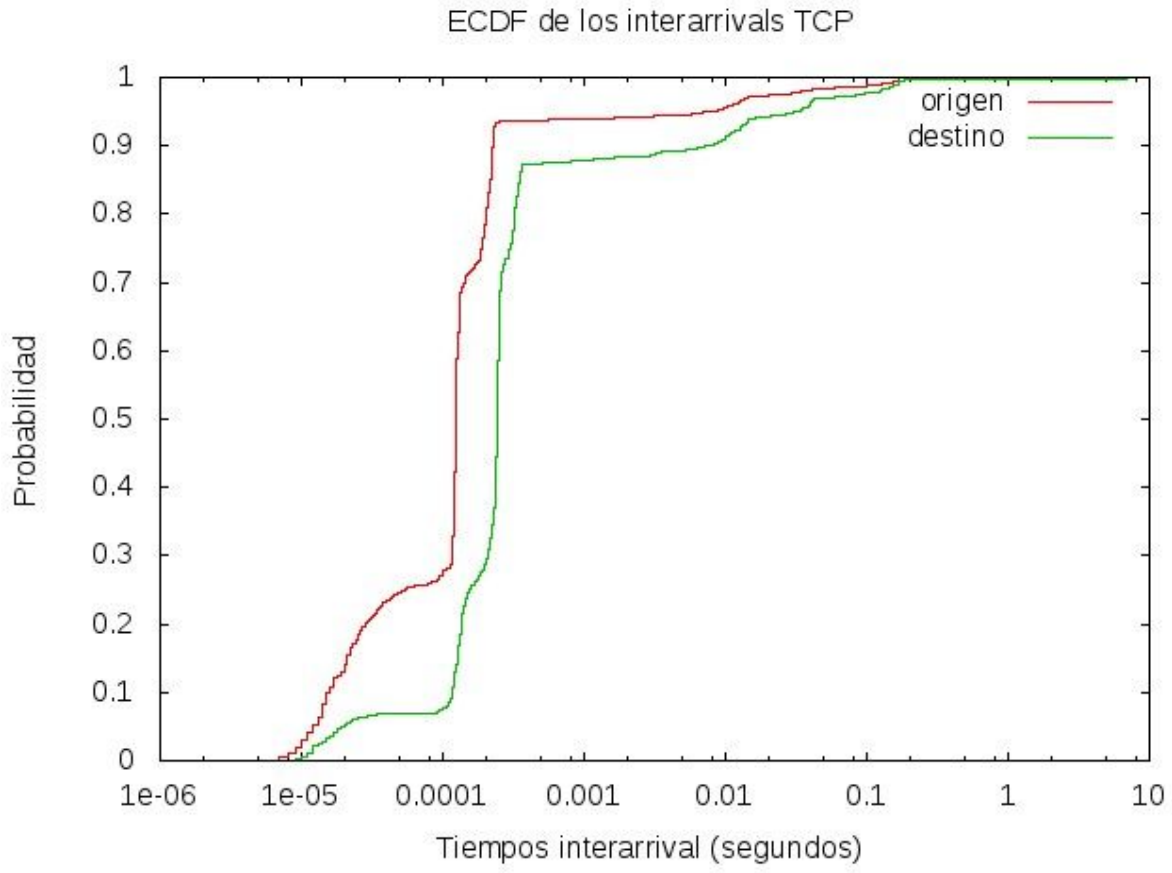
En este apartado se nos pide, utilizando que los paquetes DNS son aquellos que utilizan el puerto 53 de UDP en origen o destino, representar el tamaño a nivel tres de los paquetes. La gráfica que se muestra a continuación representa estos datos. La línea roja describe el flujo de salida ya que se ha filtrado para que tan solo se tengan en cuenta los paquetes cuyo puerto de origen UDP coincida con el valor 53 anterior; mientras que la verde se corresponde con el flujo de entrada al ser paquetes con el puerto de destino dado.



Con esta función de distribución acumulada somos capaces de ver fácilmente el comportamiento de la red en esta situación. Se aprecia que los paquetes salientes son en general cortos y los de entrada son aún más cortos. Los paquetes son pequeños con respecto a los HTTP anteriores porque encapsulan UDP que, al no llevar control de errores ni otras comprobaciones, es más corto que TCP. Los paquetes salientes (rojo) son más grandes que los entrantes (verde) porque los salientes van realizando consultas a los distintos servidores hasta llegar al adecuado y la respuesta de este, será simplemente la consulta resuelta. También usan la comunicación cliente-servidor como los apartados anteriores. Observando la gráfica, podemos intuir que las ECDFs se asemejan a una distribución normal aunque no lo sean.

2.6 ECDF Tiempo de Llegada TCP

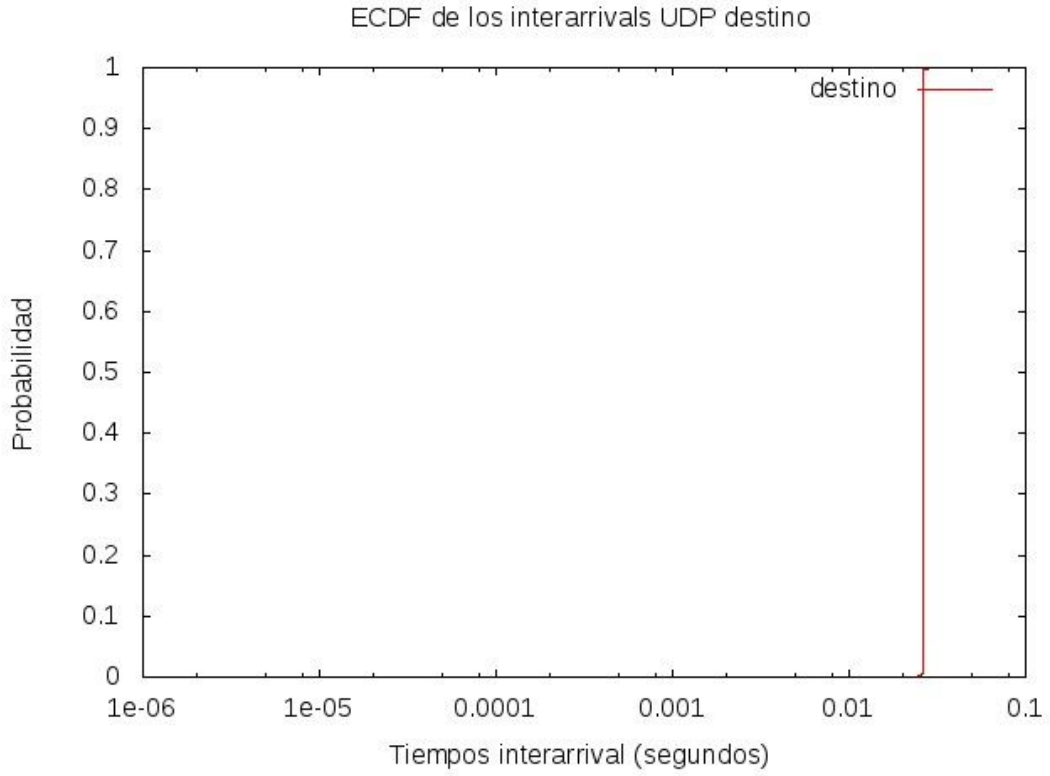
En este apartado se nos pide, utilizando la dirección IP proporcionada por el generador (53.59.245.211), representar el tiempo de llegada de los paquetes. La gráfica que se muestra a continuación representa estos datos. La línea roja describe el flujo de salida ya que se ha filtrado para que tan solo se tengan en cuenta los paquetes cuya dirección IP origen coincida con la dada; mientras que la verde se corresponde con el flujo de entrada al ser paquetes con la dirección de destino proporcionada para la traza de nuestra pareja.



Gracias a esta ECDF podemos observar con facilidad el comportamiento de la red en esta situación. Tanto en este caso como para el siguiente apartado hemos usado la escala logarítmica con respecto al eje x para ver más claros los resultados al tratarse de tiempos. Los datos obtenidos son los esperados ya que el tiempo entre llegadas de ambos flujos es muy parecido puesto que un paquete tarda un tiempo parecido en ser enviado y en ser devuelto. Las pequeñas variaciones entre paquetes salientes y entrantes pueden haber sido producidas por pequeños retardos así como las distintas probabilidades de los paquetes de un mismo flujo pueden ser debidos a las comprobaciones de errores que realizan aquellos con el protocolo TCP.

2.7 ECDF Tiempo de Llegada UDP

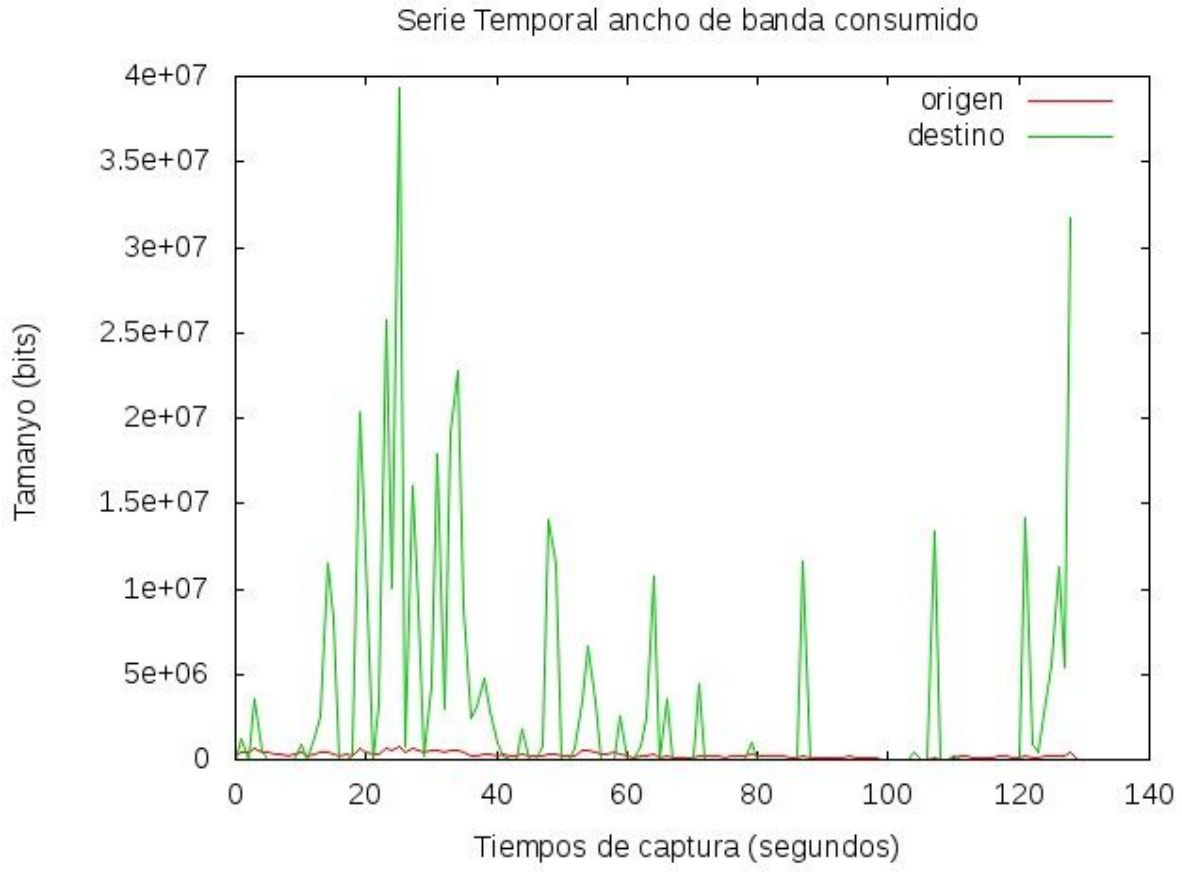
En este apartado se nos pide, utilizando el puerto proporcionada por el generador (1944), representar el tiempo de llegada de los paquetes. La gráfica que se muestra a continuación representa estos datos. La línea roja describe el flujo de entrada ya que se ha filtrado para que tan solo se tengan en cuenta los paquetes cuyo puerto destino coincida con el dado. En este punto debemos mencionar que no se muestran aquellos cuyo puerto origen sea el dado porque no existían en la traza que se nos proporciona. Esto podría asemejarse a una situación real en la cual un servicio de streaming solo se manda información en un único sentido.



Gracias a este ECDF podemos ver las diferencias fundamentales entre los protocolos TCP y UDP. UDP es un protocolo no orientado a la conexión, por tanto, el destinatario de estos paquetes no llegará ni a conocer el emisor de los mismos. Debido a su no conexión, este protocolo no manda ningún tipo de respuesta al emisor una vez se ha recibido el paquete luego el tiempo entre llegadas va a ser siempre constante. Distinto es TCP ya que es un protocolo orientado a la conexión y que realiza la confirmación de si le ha llegado un paquete además de ciertas comprobaciones de integridad de los datos. También posibilita por tanto, el reenvío de un paquete corrupto haciendo que sus tiempos entre llegadas como se ha visto en el apartado anterior sean menos constantes. Cabe destacar que la función de distribución obtenida en este apartado se trata de una distribución determinista a el valor aproximado de 0,02 segundos en el eje de abscisas.

2.8 Ancho de Banda Nivel 2

En este apartado se nos pide una figura del caudal a nivel 2 en bits/segundo. La gráfica que se muestra a continuación representa estos datos asumiendo que la dirección ethernet a considerar es la generada por nuestra traza. La línea roja describe el flujo de salida mientras que la verde describe el flujo de entrada.



Analizando la serie temporal que hemos generado, se puede observar como las peticiones que realiza el usuario (la línea roja) tienen un tamaño pequeño y relativamente constante en comparación con las respuestas recibidas (línea verde) que son más grandes por lo general y mucho más irregulares. Lo que aquí se pretende mostrar es que aquellos datos que recibe el usuario van a ser muchos más de los que manda. El usuario por lo general manda unas ciertas peticiones y espera recibir datos que en función de la petición serán más o menos grandes pero casi siempre y lógicamente, resultan más grandes que las peticiones.

3. Conclusiones

Para sacar las conclusiones de esta práctica, primero vamos a realizar un análisis del procedimiento de trabajo realizado para las mismas.

El principal objetivo de la práctica era convertirse en un gestor de red que analice ciertos parámetros y métricas de la misma para entender en qué falla y en qué se puede mejorar una red. Primero tuvimos que simular una traza usual dada mediante el generadorPCAP. En segundo lugar, tuvimos que realizar un trabajo previo: tuvimos que aprender a usar bash y awk para realizar un script que nos permitiese sacar la información que queremos analizar. Este trabajo previo sin duda fue el más laborioso y costoso puesto que empezar de cero a hacer scripts no resulta sencillo. Por último, faltaba lo que realmente era el objetivo de la práctica, realizar un estudio detallado de nuestra red gracias a las métricas y estadísticas que habíamos obtenido con nuestro script de la traza generada previamente. Para realizar este último paso, se dispone de esta memoria a la que se le adjuntan con frecuencia gráficas para que el informe sobre nuestra red sea detallado, técnico y basado en simulaciones realizadas por nuestro script.

No ha sido nada fácil tampoco la realización de este último paso ya que si bien nuestra intención es ser auténticos gestores de redes, nuestros conocimientos sobre las mismas son, de momento, limitados. En clase se ha visto con detalle las capas de red y de enlace pero los protocolos TCP y UDP aún siguen entrañando cierto misterio para nosotros, al igual que nos ocurre con los paquetes HTTP y DNS; sin embargo gracias a la ayuda de las clases prácticas y del “pozo de información sin fondo” que es Google hemos conseguido aprender un poco más sobre su funcionamiento.

A lo largo de la memoria, se han discutido aspectos como las diferencias entre los protocolos TCP y UDP, se han visto reflejadas en las gráficas características de las comunicaciones cliente-servidor, se ha obtenido más información sobre los paquetes HTTP Y DNS, se ha comprobado cuáles son los protocolos más usados en una traza usual,...

En definitiva, ha sido una práctica muy completa y de arranque perfecto para nuestra formación como gestores de redes.