Práctica 3: Análisis de Tráfico

Introducción

El análisis de tráfico de red es un elemento clave dentro de los departamentos de TI de cualquier empresa. Por ello, es necesario abordarlo desde distintos contextos: monitorización, diagnóstico de problemas de red, análisis forense, seguridad, etc.
En esta práctica nos adentraremos en el uso del aplicativo tshark, que nos permitirá capturar tráfico y/o procesarlo en función de parámetros concretos.

Aplicativo *tshark*

El aplicativo *tshark* es la versión basada en texto del *framework* de análisis de tráfico de red *Wireshark*. Gracias a él, se podrá realizar el análisis en vivo de una red de comunicaciones o una captura estática de tráfico de red, clasificando paquete a paquete por su protocolo de comunicaciones y evaluando cada uno de los detalles pertenecientes a dicho protocolo. Además, se podrá realizar tratamiento estadístico de los datos capturados con el fin de obtener resultados más avanzados. Este tipo de análisis requiere que la interfaz de red donde se capture la información esté configurada en modo promiscuo, siendo necesario, en la mayoría de los casos, permisos de super-usuario para su puesta en marcha.

El tshark no suele venir instalado por defecto en las distribuciones de Linux. No obstante, para instalarlo en cualquier ordenador con un sistema basado en Debian, como por ejemplo Ubuntu, se puede ejecutar de la siguiente manera:

sudo apt-get install tshark

Toda la información relacionada con los diferentes parámetros de entrada del aplicativo puede consultarse desde un sistema operativo GNU Linux de la siguiente manera:

man tshark

IMPORTANTE! El aplicativo tshark YA se encuentra instalado en los equipos del laboratorio por lo que NO será necesario instalar ninguna herramienta en ellos.

Ejercicios

Para responder a las preguntas de evaluación citadas a continuación se deberá hacer uso **obligatorio** del aplicativo tshark, por lo que no serán admitidas dentro de la memoria justificativa otro tipo de aplicativos similares para el procesamiento del tráfico de red. El aplicativo tshark se ejecutará pasando como argumento obligatorio un fichero PCAP generado previamente a partir del script proporcionado en la página de Moodle de la asignatura. No obstante, para la realización de los ejercicios planteados se podrán utilizar cualquier comando de tratamiento de datos proporcionado por el sistema operativo GNU Linux, como por ejemplo awk, sed, tr, sort, wc -l etc.

Se deberá entregar una memoria justificativa (en formato PDF). Esta memoria será el elemento a evaluar y por tanto es OBLIGATORIA su entrega. Esta memoria debe incluir respuesta de manera detallada e indicando el comando *tshark* necesario MÁS los comandos del sistema operativo/scripts/programas-propios para obtener la respuesta/representaciones/gráficas de las siguientes preguntas/visualizaciones:

Operativa básica (0.5+0.5+0.75 puntos)

- ¿ Cuántos paquetes tiene la traza de tráfico de red ?
- ¿ Cuál es el tamaño medio del paquete ?
- ¿ Cuál es la tasa media en kbps (bits por segundo) del enlace que monitoriza la traza ? ¿ y en pps (paquetes por segundo) ?

Operativa basada en filtrado (0.5+0.5+0.75 puntos)

Escoja una dirección MAC de la traza.

- Quédese solo con el tráfico destinado a esa MAC elegida. ¿Cuál es la tasa media (en kb/s) del tráfico de red?
- Quédese solo con el tráfico originado desde la MAC en estudio ¿Cuál es la tasa media (en kb/s) del tráfico de red?
- ¿ Cuántos paquetes hay que no tengan esta MAC en origen ni en destino ?

Operativa de protocolo a nivel de paquete (1+1+1+1.5+1 puntos)

Filtre la traza proporcionada en función del siguiente esquema y responda a las siguientes cuestiones:

- Paquete de inicio: Número de pareja * 10 (en caso de ser la pareja 10, la traza filtrada deberá comenzar a partir del paquete número 100).
- Paquete de fin: Paquete de inicio + 500 (en caso de ser la pareja 10, la traza filtrada deberá acabar en el paquete 600).

- ¿ Cuántos paquetes de la traza son IP ? ¿ Y ARP ? ¿ Hay otro tipo de protocolos ? En caso afirmativo, ¿ sabrías decir de qué protocolos son ?
- De los paquetes IP, ¿ Cuántos paquetes TCP hay en la traza ? ¿ y UDP ? ¿ e ICMP ? ¿ hay paquetes de otros protocolos ? En caso afirmativo, ¿ sabrías decir qué tipo de paquetes son ?
- Basándose en la dirección IP origen y destino así como del puerto origen y destino del tercer paquete de la traza. ¿ Cuál sería el número de paquetes y la tasa media en bps (bits por segundo) del flujo UDP al cual pertenece el paquete previamente descrito ?
- Graficar la función ECDF (*Empirical Cumulative Distribution Function*) para el tamaño de todos los paquetes, así como sólo aquellos pertenecientes al protocolo TCP y sólo aquellos pertenecientes al protocolo UDP. ¿ Se observa alguna moda en cualquiera de los casos ? ¿ Qué porcentaje de paquetes son mayores de 400 bytes en cualquiera de los casos ?
- ¿ Cuáles son los 5 puertos (destino y origen) más populares tanto en número de paquetes como en número de bytes ? ¿ A qué se debe la diferencia de popularidad en paquetes y en bytes ?

Implementación (1 punto)

- Implementación coherente y óptima de las instrucciones tshark y otras, una vez se hayan respondido correctamente el resto de cuestiones.

Fechas

La entrega deberá hacerse antes de las 00:00 del día límite establecido en la página de Moodle de la asignatura.

Examen

Se realizará un examen INDIVIDUAL el día indicado en la página de Moodle de la asignatura. De manera similar a evaluaciones anteriores, el examen consistirá en unas cuestiones sobre la ejecución concreta de la práctica por parte de cada pareja de prácticas.
La calificación del examen será APTO/NO APTO. En caso ser NO APTO, la calificación de la práctica para ese miembro de la pareja será de 0 (cero).

TSHARK

A continuación presentamos la forma de invocar *tshark* desde la consola de comandos para conseguir obtener distintos parámetros a partir de capturas de tráfico. Para ello, distinguiremos entre características de los paquetes (p.ej., momento de captura o bytes capturados) y campos de

cabeceras (p.ej., direcciones IP o puertos), y también veremos cómo filtrar el tráfico para que los resultados se restrinjan utilizando filtros de visualización. En general, para obtener una serie de campos de datos a partir de una captura de tráfico debemos utilizar la llamada:

```
tshark -r <nombre del archivo pcap> -T fields [-e <nombre de campo>]
```

A continuación desglosaremos algunas de las opciones de campos que se pueden extraer. Como veremos, los nombres de los campos se corresponden con los que se utilizan en los filtros de visualización de Wireshark.

OPCIONES PARA EXTRACCIÓN DE CARACTERÍSTICAS DE PAQUETES

tshark permite extraer distintas características de los paquetes capturados, siendo las más interesantes para nuestros intereses las siguientes:

Nombre del campo	**Significado**
frame.len	Longitud original del paquete
frame.number	Número o índice de paquete capturado.
frame.time_delta	Tiempo transcurrido desde el anterior paquete capturado
frame.time_delta_displayed	Tiempo transcurrido desde el anterior paquete mostrado
frame.time_epoch	Tiempo de captura del paquete (formato EPOCH)
frame.time_relative	Tiempo de captura del paquete (desde el comienzo)

Por ejemplo, para obtener los tiempos de captura en formato EPOCH y tamaños de los paquetes de una traza, se puede utilizar el comando:

```
tshark -r <nombre del archivo pcap> -T fields -e frame.time_epoch -e frame.len
```

OPCIONES PARA EXTRACCIÓN DE CAMPOS DE CABECERAS DE LOS PAQUETES

Por otro lado, *tshark* también permite extraer los campos de las cabeceras de los distintos protocolos presentes en el paquete. A continuación se presenta un resumen (por niveles) de algunos campos de especial relevancia:

Nivel	**Nombre del campo**	**Significado**
Ethernet	eth.addr	Dirección MAC (destino y/u origen)
	eth.dst	Dirección MAC de destino
	eth.src	Dirección MAC de origen
	eth.type	Tipo Ethernet (protocolo encapsulado)
IP	ip.addr	Dirección IP (destino y/u origen)
	ip.dst	Dirección IP de destino
	ip.src	Dirección IP de origen
	ip.proto	Protocolo encasulado en IP
TCP	tcp.dstport	Puerto TCP de destino
	tcp.srcport	Puerto TCP de origen
UDP	udp.dstport	Puerto UDP de destino
	udp.srcport	Puerto UDP de origen

Por ejemplo, para obtener una salida de texto que incluya los tiempos de captura en formato EPOCH y puerto destino UDP se podría usar la siguiente invocación:

```
tshark -r <nombre del archivo pcap> -T fields -e frame.time_epoch -e udp.dstport
```

Es importante señalar que los registros que proporciona la salida de *tshark* si se usan este tipo de llamadas pueden tener un número variable de campos. Si solicitamos que se muestren campos correspondientes a la cabecera de un cierto protocolo, los paquetes que no incluyan dicho protocolo generarán registros con campos vacíos. Como veremos en el apartado siguiente, esto se puede resolver utilizando filtros de visualización que fuercen que el análisis se ejecute únicamente sobre paquetes que cumplan características como incluir un cierto protocolo.

Por otro lado, también es posible obtener una disección completa de los paquetes (similar a la proporcionada en la interfaz de Wireshark) utilizando *tshark*. Para ello, debemos invocar el programa usando la siguiente llamada:

```
tshark -r <nombre del archivo pcap> -T text -V
```

APLICACIÓN DE FILTROS DE VISUALIZACIÓN

Al igual que en la interfaz gráfica de Wireshark, *tshark* permite la introducción de filtros de visualización para discriminar el tráfico analizado. Para ello, debemos utilizar la siguiente sintaxis (nótese el uso de apóstrofes para delimitar el filtro):

```
tshark -r <nombre del archivo pcap> -Y '<filtro de visualización>'
```

De esta forma, podemos centrarnos en paquetes que cumplan un conjunto de características: por ejemplo, para obtener el tiempo en el que fueron capturados paquetes TCP con puerto origen 5060 podríamos usar la invocación:

```
tshark -r <nombre del archivo pcap> -T fields -e frame.time_epoch -Y 'tcp.dstport eq 5060'
```

Como comentamos antes, esto permite evitar problemas asociados en el caso de solicitar campos que sólo estén presentes en paquetes correspondientes a protocolos particulares. Por ejemplo, en el caso del apartado anterior en el que obteníamos el puerto destino UDP, podríamos suprimir la salida de registros que no tengan un valor vacío en este campo modificando la invocación que indicábamos por:

```
tshark -r <nombre del archivo pcap> -T fields -e frame.time_epoch -e udp.dstport -Y 'udp'
```

IMPORTANTE, la sintaxis de filtros de visualización de Wireshark y tshark es la misma, así que ante la duda, la interfaz gráfica nos puede ayudar a comprobar que los filtros que estamos usando tienen una estructura correcta.

Representación Gráfica de Datos

Uno de los ejercicios propuestos en la práctica pide realizar el graficado de funciones ECDF (*Empirical Cumulative Distribution Function*). Para ello, se hará uso del aplicativo *gnuplot* instalado en los laboratorios. Además, se proporcionará en la página Moodle de la asignatura un *script bash* encargado de realizar la llamada correspondiente al aplicativo *gnuplot*. Dicho *script* llamado plot_cdf.sh recibe como argumentos el nombre del fichero de datos así coma los valores de las etiquetas del título y del eje X para mostrar en la gráfica y generar un archivo de extensión PNG con la gráfica dibujada.

Ejemplo: ./plot_cdf.sh fichero_de_datos.dat "Titulo" "Eje X"

Por último, con el objetivo de aplicar el concepto de funciones ECDF se explicar en el laboratorio su significado y utilidad. Se recomienda leer el siguiente enlace: **https://en.wikipedia.org/wiki/Empirical_distribution_function**

IMPORTANTE! Es OBLIGATORIA la utilización del script de graficado plot_cdf.sh quedando PROHIBIDA la utilización de cualquier otra herramienta o aplicativo encargado del graficado de la función ECDF.

Generación de la traza de prueba

La traza de tráfico de red que se deberá usar para de dicha práctica será **única** para cada pareja. Para ello, se utilizará el número de DNI del primer miembro de la pareja. Dicho número se pasará coma argumento obligatorio al programa que se proporciona en la página Moodle de la asignatura (denominado generateTrace) y generará la traza PCAP que se deberá utilizar para la práctica.

Por ejemplo, si el número de pareja es el 9 y el DNI es 12345678 ejecutaremos:

chmod +x generateTrace
./generateTrace 9 12345678

Este comando producirá una traza con nombre, en este ejemplo concreto, practica3_9_12345678.pcap que deberá ser utilizada para la realización de la práctica.

IMPORTANTE! Es FUNDAMENTAL que el NÚMERO DE DNI aparezca en la MEMORIA que se adjunte.