Práctica 3, sesiones 6/7/8: Monitorización

Objetivos de la práctica*

- Introducir la monitorización de redes de comunicaciones
- Introducir un nuevo protocolo de red: *VLAN*
- Utilizar herramientas de *scripting* para el análisis de protocolos

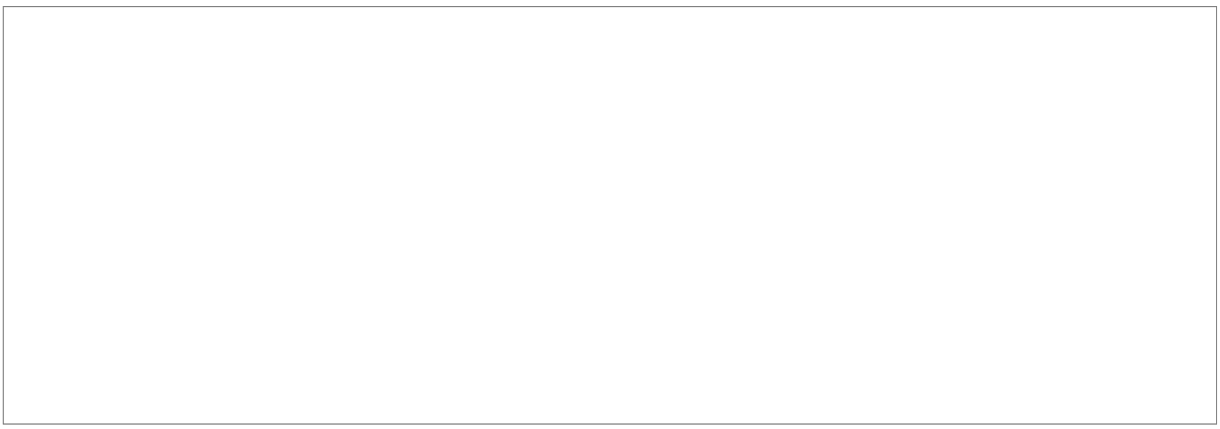
Introducción

La monitorización de una red consiste en el análisis de rendimiento y estado, con el fin de detectar problemas, ineficiencias o ataques, que de otra forma no podrían ser detectados o mitigados. Un análisis exhaustivo del estado de red puede abarcar tanto problemas de la propia red como problemas de los equipos y aplicaciones que la componen.
La monitorización de red se puede dividir en dos tipos fundamentales:

- La **monitorización activa** consiste en transmitir determinados paquetes por la red con el objetivo de medir ciertos parámetros de la misma, (ancho de banda disponible, retardos, *jitter*, etc). No obstante, este tipo de monitorización puede afectar al comportamiento de la red.
- La **monitorización pasiva**, por otro lado, analiza el tráfico que está atravesando la red ofreciendo métricas complementarias a las activas (información de flujos, etc.). Este tipo de monitorización, a diferencia de la activa, no afecta al estado de la red.

El objetivo de esta práctica es que el alumno se ponga en el papel de un gestor de red, obteniendo diversas métricas pasivas analizando tráfico real de red. Su trabajo consistirá en estudiar el funcionamiento y rendimiento de la red que gestiona. Para lograrlo, el gestor de red se propone desarrollar las herramientas necesarias para realizar dicha tarea y utilizarlas para realizar un informe sobre aspectos concretos de la red que le han pedido.

La primera tarea que debería realizar el gestor es capturar en tiempo real el tráfico a estudiar, para obtener un conjunto de datos que permita realizar análisis posteriores. En nuestro caso, se facilita al estudiante un generador de trazas .pcap, que genera una traza para cada pareja y grupo de prácticas, de forma que cada pareja pueda trabajar sobre datos de tráfico diferentes. La traza generada simula el tráfico entre 2 routers que transportan información de audio, web y otros servicios entre los diferentes elementos que interconectan (ver siguiente figura).



En las redes de comunicaciones, es común querer diferenciar tráfico de diferentes tipos, por ejemplo, para establecer diferentes políticas de calidad (en términos de ancho de banda, *jitter*, latencia...) o para distinguir tráfico de diferentes clientes en un medio compartido. Una de las formas para conseguirlo es utilizando etiquetas de VLAN (IEEE 802.1Q). En Ethernet, la etiqueta de VLAN tiene como *ethertype* el **0x8100**, y **el administrador de red debe ser consciente de que algunos paquetes de su red pueden tener una o más etiquetas VLAN**. Aunque no es una representación literal del estándar IEEE 802.1Q, puede utilizar la siguiente imagen como descripción de VLAN para las prácticas. En el campo "Type" de este protocolo, nos encontraremos el siguiente EtherType (que bien podría ser IP o ARP).

Una vez el administrador de red dispone de una serie de herramientas que entienden los protocolos más comunes de red (Ethernet, VLAN, IP, TCP y UDP), es posible comenzar a estudiar el tráfico. Para ello, el administrador de red extrae las estadísticas y métricas más básicas de la traza. Recordemos, que el objetivo de esta práctica no es la codificación de las herramientas, sino la realización de los análisis e interpretación de los resultados. Por este motivo, el estudiante deberá realizar una memoria/informe en la que describa el comportamiento de la red, así como la interpretación de los datos obtenidos.

El informe presentado deberá incluir:

- La popularidad por protocolos en bytes y en paquetes.
- Top 10 de direcciones IP y puertos (TCP o UDP) activos (en cantidad de bytes y de paquetes).
- Ancho de banda consumido.
- Una caracterización estadística de algunas conversaciones de interés. Para ello, se utilizarán una serie de funciones de distribución acumulada empírica (ECDFs), de acuerdo a diferentes protocolos y sentidos del tráfico.

Recursos proporcionados

- **generadorPCAP**: Para la realización de la práctica se proporciona a los estudiantes una aplicación que generará una traza pcap con características únicas para cada

pareja y grupo. **El generador proporcionará por pantalla una serie de información que el estudiante deberá utilizar para el análisis posterior** del fichero pcap. En caso de que la pareja no utilice los datos indicados por el generador o utilice la traza de otra pareja/grupo, se **considerará que se trata de una copia** y la practica **será calificada con 0 puntos**. Para facilitar la tarea a los estudiantes, se proporcionan 2 ejecutables diferentes (para las arquitecturas x32 y x64), de forma que puedan generar la traza tanto en la máquina virtual proporcionada (x32) como un otro sistema Linux de 64 bit. La aplicación debe ser ejecutada de la siguiente forma:

./generadorPCAPx32 <numero grupo> <numero pareja> <traza.pcap>

- **ejemploGNUplot.gp**: Se facilita un *script* de ejemplo para pintar datos con gnuplot (si se desea profundizar en el uso de la herramienta gnuplot, puede visitar la siguiente página web: http://www.duke.edu/~hpgavin/gnuplot.html). El *script* asume un archivo de entrada llamado *salida.txt,* que deberá ser la salida de algún programa del estudiante. Se espera que dicho fichero contenga dos columnas, una muestra por fila, de forma que se pinta la primera coordenada contra la segunda. Tras esto, el script imprime la gráfica por la terminal, y se generan 3 ficheros con la imagen de la misma en distintos formatos. Para ejecutar el script basta con:
  - Dar permisos de ejecución al fichero: *chmod +x ejemploGNUplot.gp*
  - Ejecutar desde una terminal: *./ejemploGNUplot.gp*
- **crearCDF.c**: **Código inicial** para una aplicación que calcula una ECDF. La versión facilitada solo cuenta el número de muestras y las ordena. Estos son pasos iniciales para el cálculo de una ECDF, siendo trabajo del estudiante completar el programa o bien hacerlo mediante algún script.
- **Guía de análisis de red con *tshark* y *shell scripting***: Dado que no se plantea codificar un analizador de tráfico, se parte del programa *tshark*, que puede considerarse una versión en modo texto de Wireshark que permite trabajar por línea de comandos. Dicho programa, junto con distintos comandos típicos en una *shell* de Linux, permitirán realizar los análisis que se plantean.

Ejercicios a realizar

Implementación

Para realizar el informe/memoria, el estudiante deberá realizar una serie de *scripts* basados en *tshark*, *shell* y *awk*, que permitan realizar todos los análisis necesarios. Para guiarse durante el proceso de implementación de estos *scripts*, puede consultar la guía proporcionada.

Por sencillez, se recomienda que el alumno asuma que se tiene una traza con el tráfico ya capturado, y que realice varias invocaciones de la herramienta *tshark* para obtener resultados parciales que pueda luego utilizar para generar los resultados a incluir en la memoria. Se espera que finalmente se tenga un *script* que incorpore todas las llamadas necesarias para obtener los resultados que se incluirán en el informe a entregar.

Si la máquina virtual no tiene instalado el *tshark*, puede hacerlo invocando desde la línea de comandos:

sudo apt-get install tshark

Funcionalidad de los *scripts* desarrollados

Los scripts a desarrollar deberán:

1. Obtener el porcentaje de paquetes IP y no IP, y dentro de los que sean IP {UDP, TCP u OTROS (Ni TCP ni UDP)} (punto 1 de la memoria). **Se entenderá como paquete IP todo aquel cuyo tipo Ethernet es IP (0x0800), o bien el tipo Ethernet es VLAN (0x8100) y el tipo VLAN es IP (0x0800).**
2. Calcular el "top 10" tanto en bytes como paquetes de:
   - **Direcciones IP**
   - **Puertos** (considerando por separado TCP y UDP).
   - **En ambos casos, se hará distinción de si es dirección/puerto origen o destino.**
3. Almacenar en un archivo la ECDF de los tamaños de los paquetes leídos.
4. Almacenar en un archivo la serie temporal del caudal/*throughput*/tasa/ancho de banda por cada sentido, con una granularidad de 1 segundo y **medido a nivel 2 en bits por segundo (b/s)**. **OJO: filtre a nivel Ethernet para separar el tráfico en cada sentido.**
5. Almacenar en un archivo la ECDF de los tiempos entre llegadas de los paquetes de los flujos indicados por el generador de trazas en ambos sentidos (entendamos flujo por aquellos paquetes que cumplen unas determinadas condiciones).
6. Obtener representaciones gráficas adecuadas para los resultados obtenidos en los puntos 2-5.

Memoria

La memoria o informe a entregar debe estar escrito en un lenguaje formal y correcto. **Deberá incluir una portada, una breve introducción, así como un apartado de conclusiones. Los análisis realizados deberán mostrar, explicar (o al menos intuir) y razonar cada una de las medidas solicitadas sobre la traza proporcionada. A la hora de evaluar la práctica, no se admitirán resultados que no incluyan estas explicaciones y razonamientos.** Por otro lado, es importante también indicar la razón o motivación que el estudiante cree que hay detrás de la estadística solicitada y por qué se considera que puede resultar de interés en tareas de monitorización.

Según lo mencionado previamente, la memoria debe incluir:

1. Porcentajes de paquetes (puede incluir una captura de pantalla) (punto 1 de los requisitos):
   - IP y NO IP (entendemos como **NO-IP** aquellos paquetes que no son ni **ETH|IP** ni **ETH|VLAN|IP**)

- UDP, TCP, OTROS sobre los que son IP (igualmente entienda, un paquete IP como aquel que cumpla la pila **ETH|IP** o **ETH|VLAN|IP**).
- **Indique en todos los casos la expresión de filtro utilizada.**

2. Top 10 de direcciones IP activas (en bytes y paquetes, y por sentido) y top 10 de puertos (en bytes y paquetes, y por sentido) (una captura de pantalla puede ser suficiente).
3. **ECDF** de los tamaños a nivel 2 de los paquetes de la traza (una por sentido, utilice la dirección MAC proporcionada por el generador).
4. **ECDF** de los tamaños a nivel **3** de los paquetes HTTP de la traza (una por sentido a nivel 4). Entenderemos como HTTP todos aquellos paquetes que usen el puerto 80 de TCP en origen o destino.
5. **ECDF** de los tamaños a nivel **3** de los paquetes DNS de la traza (una por sentido a nivel 4). Entenderemos como DNS todos aquellos paquetes que usen el puerto 53 de UDP en origen o destino.
6. **ECDF** de los tiempos entre llegadas del flujo TCP indicado por el generador de la traza (una por sentido a nivel 4).
7. **ECDF** de los tiempos entre llegadas del flujo UDP indicado por el generador de la traza (una por sentido a nivel 4).
8. Figura (o figuras) que muestre(n) el caudal/*throughput*/tasa/ancho de banda a **nivel 2 en bits por segundo (b/s)** y por sentido (asuma que la dirección Ethernet origen o destino es la indicada por el generador de trazas). Los segundos sin tráfico deben representarse a cero.
9. **Todos los resultados obtenidos deben ser explicados e interpretados por los miembros de la pareja, y quedar reflejados en la memoria.**

   Nota: Entendemos "sentido por nivel 4", como aquel flujo que tiene como origen y destino unos puertos determinados.

Entrega

Denomine a los archivos de entrega **practica3_analisis.sh (script principal de análisis) y practica3_memoria.pdf**. Añada un archivo **leeme.txt** que incluya los nombres de los autores, comentarios que se quieran transmitir al profesor (mejoras que podrían hacerse a la práctica, etc.) y descripción de los scripts auxiliares que se hayan implementado.

Comprima en un **zip** todo lo que vaya a entregar y llámelo **practica3_YYYY_PXX.zip**, donde YYYY es el grupo al que pertenece (1301,1302,etc), y XX (y solo XX) es el número de pareja (con dos dígitos).
Por ejemplo mediante:

$ zip practica3_1301_P01.zip practica3_1301_P01/*

(Asumiendo que se quieren entregar todos los archivos de la carpeta con el mismo nombre y es la pareja 1 del grupo 1301).

Criterios de evaluación

Ejercicio: Entrega el día 16 de noviembre hasta las 23:55h. En caso de fallo de Moodle recuerde que puede utilizar el correo, pero no se admitirán entregas fuera de plazo. **NO ENTREGUE EL GENERADOR DE TRAZAS NI LA TRAZA GENERADA. EL ARCHIVO ZIP SERÁ DEMASIADO GRANDE Y NO PODRÁ SUBIRLO A MOODLE.**

- **Scripting (50%):**
  - Cálculo de porcentajes por protocolos: **5%**
  - Obtención del top de puertos: **5%**
  - Obtención del top de direcciones IP: **5%**
  - Cálculo del caudal/*throughput*/tasa/ancho de banda por sentido: **10%**
  - Obtención de la ECDF del tamaño de paquetes: **15%:**
    - **5%, realizada sobre toda la captura**
    - **5%, realizada sobre los paquetes HTTP**
    - **5%, realizada sobre los paquetes DNS**
  - Obtención de la ECDF de los *interarrivals*/intervalos de los flujos indicados por el generador de PCAP: **10%**
    - **5%, correspondiente al flujo UDP**
    - **5%, correspondiente al flujo TCP**
- **Memoria (50%):**
  - Porcentajes por protocolos: **5%**
  - Top de puertos: **5%**
  - Top de direcciones IP: **5%**
  - Series temporales del caudal/*throughput*/tasa/ancho de banda por sentido: **10%**
  - ECDFs del tamaño de paquetes: **15%:**
    - **5%, realizada sobre toda la captura**
    - **5%, realizada sobre los paquetes HTTP**
    - **5%, realizada sobre los paquetes DNS**
  - ECDFs de los *interarrivals*/intervalos de los flujos indicados por el generador de PCAP: **10%**
    - **5%, correspondiente al flujo UDP**
    - **5%, correspondiente al flujo TCP**
- **IMPORTANTE: los resultados presentados en la memoria que no incluyan un razonamiento, explicación o motivación no serán contabilizados. Esto es, su valoración será 0%.**

Cuestionario: El día 27 de octubre se realizará un breve cuestionario individual en Moodle. El objetivo del cuestionario será evaluar que los estudiantes hayan afianzado los objetivos de la práctica. Se considera que para entonces debe haberse realizado (o llevar muy avanzados) los tres primeros puntos del ejercicio (porcentajes de protocolos, top puertos y direcciones IP, y *throughput*), y que se manejan los contenidos del manual de análisis proporcionado. De este modo, se supervisará que el enfoque del informe sea el más correcto para poder atajar lo antes posible problemas que puedan surgir durante el desarrollo de la práctica. En cualquier caso, el cuestionario es optativo y la asistencia a la sesión práctica no será obligatoria.

Control individual: El control de la práctica se realizará el día 17 de noviembre. Sea puntual, pues comenzamos a "y 5".

---

* Como en prácticas anteriores este enunciado es descriptivo de los requisitos de las entregas y criterios de evaluación. No es por tanto necesariamente una guía "HOW-TO" para su desarrollo, en el laboratorio se explicará, concretará y se determinará cómo realizar el trabajo.

Última modificación: viernes, 20 de octubre de 2017, 17:42