Redes de Computadoras

Obligatorio 1 - 2019

Facultad de Ingeniería
Instituto de Computación
Departamento de Arquitectura de Sistemas

Nota previa - IMPORTANTE

Se debe cumplir íntegramente el "Reglamento del Instituto de Computación ante Instancias de No Individualidad en los Laboratorios", disponible en el EVA.

En particular está prohibido utilizar documentación de otros estudiantes, de otros años, de cualquier índole, o hacer público código a través de cualquier medio (EVA, news, correo, papeles sobre la mesa, etc.).

Introducción

Objetivo del Trabajo

Familiarizarse con conceptos básicos sobre redes e Internet y manejar herramientas para diagnóstico y *debug*. Asimismo, esta tarea intenta que el estudiante se plantee interrogantes e investigue sobre temas que serán abordados durante el curso.

Herramientas

La tarea se puede desarrollar en cualquiera entorno de trabajo, dependiendo de los permisos requeridos por las herramientas necesarias que son las siguientes:

- `ping`
- `wireshark`
- `traceroute (equivalente en Windows: tracert)`
- `dig`

En caso de requerir permisos de root, se recomienda desarrollar la actividad en una máquina virtual.

Forma de entrega

Una clara, concisa y descriptiva documentación es clave para comprender el trabajo realizado. La entrega de la tarea consiste en un único archivo `obligatorio1GrupoGG.tar.gz` que deberá contener los siguientes archivos:

- Un documento llamado `Obligatorio1GrupoGG.pdf` donde se documente todo lo solicitado en la tarea. GG es el número del grupo.
- Los programas y capturas solicitados.
- Un directorio `extras` incluyendo cualquier otro archivo que considere relevante.

La entrega se realizará en el sitio del curso, en la plataforma EVA.

Fecha de entrega

Los trabajos deberán ser entregados **antes del 25/8/2019 a las 23:30 horas**. No se aceptará ningún trabajo pasada la citada fecha y hora. En particular, no se aceptarán trabajos enviados por e-mail a los docentes del curso.

Observaciones

Los programas pueden ser escritos en cualquier lenguaje, pero se recomienda utilizar algún lenguaje de scripting, que son adecuados al tipo de tarea solicitada (`shell script`, `php` o `python` por ejemplo). Los programas deberán poder ejecutarse en los PCs Linux de Facultad.

Toda vez que se pida la ejecución de un comando y una respuesta, analice dichos resultados; la ejecución del mismo, incluyendo su invocación deberá ser parte de la respuesta.

Todas las capturas solicitadas deberán ser almacenadas y entregadas en el formato pcap.

Parte A – Análisis de logs

La Universidad de la República alberga algunos mirrors de proyectos de software libre, de escala global. Se tienen mirrors tanto de sistemas operativos, lenguajes y paquetes de ofimática, entre otros. Particularmente, trabajaremos con los logs de `espejito.fder.edu.uy`.
`espejito` sirve su contenido en base al servidor web apache versión 2.4.39, y a su grupo se le asignó un archivo de log único para analizar. Se recomienda lea la documentación de apache, particularmente, del formato de los archivos de log, a efectos de programar la primer parte de la tarea.
Su archivo de log se encuentra en
`https://espejito.fder.edu.uy/redes-ob-1/<numero-grupo>.xz`

En base a la información del log usted deberá contestar a las siguientes preguntas:

(a) ¿Cuántos bytes transmitió el servidor de acuerdo al log?
(b) ¿A cuántas IP distintas sirvió?
(c) ¿A cuáles sistemas autónomos sirvió?
(d) ¿A cuántos países sirvió? ¿con qué frecuencia?
(e) ¿Cuántos bytes por protocolo? (IPv4 e IPv6)
(f) ¿Qué versiones del protocolo HTTP fueron utilizadas y con qué frecuencia?
(g) Determine la cantidad de conexiones por hora que se recibieron e indique a qué hora hubo más conexiones.
(h) Determine la cantidad de bytes transmitidos por hora e indique a qué hora hubo una mayor transmisión de bytes. ¿Coincide con el horario de la parte (g)? Explique
~~(i) Determine cuál fue el archivo de más de 1MiB (mega byte binario) que fue servido más rápidamente y a cuántos Mbps (megabits por segundo).~~
(j) En base al cálculo de ancho de banda saliente efectivo por hora utilizado por el mirror realizado en h), indique cuántos MiB (megabytes binarios) se transfirieron y el uso de ancho de banda promedio en Mbps (megabits por segundo). ¿qué enlace mínimo requiere para alojar dicho mirror?
(k) ¿Cuál fue el proyecto más activo? (proyecto opensuse, centos, fedora, tdf, etc.) Brinde un histograma con la cantidad de bytes servidos por proyecto.
(l) ¿Qué códigos de respuesta HTTP se encuentran en el archivo y cuál es su frecuencia?

Se pide:

1. ¿Qué es un mirror? ¿qué funcionalidad cumple? Explíquelo, de forma clara, utilizando sus palabras.

2. Diseñe e implemente en un lenguaje de alto nivel de su elección un analizador de logs para responder a las preguntas anteriores. Además de la implementación deberá entregar un documento de diseño donde explique claramente su solución y las estrucutras de datos definidas. Las mediciones que requieren frecuencias, deberán ser presentadas en formato de historgramas.
   El informe deberá resumir todas las respuesta, e incluir histogramas para: d), e,), f), g), h), j), k) y l)

3. Considerando que HTTP soporta descargas parciales, en particular, considerando los códigos de respuesta `200` y `206`, ¿cambia esto su definicíon de “recurso más descargado”? Explique.

4. ¿Cuáles proyectos registran descargas parciales?

5. Determine las 20 direcciones IP más activas, y su frecuencia.

Parte B - Herramientas del sistema

1. Trabajaremos ahora con las 20 direcciones IP anteriores. Utilizando el comando `ping`, indique cuál es la más próxima a usted. Fundamente.

2. Investigue el principio de funcionamiento del comando `traceroute`.

3. Utilizando `traceroute`, cuente la cantidad de saltos a cada una de dichas direcciones. ¿El criterio de proximidad que utilizó anteriormente coincide con la cantidad de saltos? Fundamente.

4. Investigue y documente brevemente el funcionamiento del sistema DNS en Internet. En este contexto, investigue y documente para que sirve y como funciona la utilidad del sistema `dig`.

5. Utilice el comando `dig` para conocer el nombre DNS de las IP anteriores.
   ¿Qué parámetro utiliza para este fin? ¿Alguna de las direcciones anteriores pertenece a usuarios domésticos? ¿Alguna de las direcciones anteriores pertenece a empresas u organismos gubernamentales conocidos? Fundamente.

Parte C - Wireshark

1. Investigue y documente para que sirve y como se utiliza la herramienta `Wireshark`.

2. Realice la captura del tráfico generado por un equipo, para la interfaz que conecta el equipo con Internet mientras realiza el acceso a través de un navegador a la página `http://ftpmirror.gnu.org/glibc/glibc-2.28.tar.gz`.
   Grabe en un archivo el tráfico capturado[1]. Analice la captura de tráfico realizada:

   a) ¿Que consultas DNS se realizan? Identifique en particular el mensaje de solicitud al servidor DNS utilizado para obtener la IP del dominio. ¿Que servidor DNS se utilizó?

   b) ¿En qué momento se produce la redirección a un mirror?

   c) ¿Qué user-agent publica su navegador al servidor? ¿Por qué es útil para el servidor web tener esta información?

   d) Identifique en su captura los segmentos TCP donde el servidor envía la página HTML principal.
      ¿Cuántos segmentos utiliza TCP para entregarle el contenido?

Notas:

- En los equipos de facultad la consulta a DNS externos esta deshabilitada, por lo tanto existen partes que deben realizarse en un equipo conectado a una red hogareña o a través de la red inalámbrica wifing utilizando el protocolo TCP.
- Tenga en cuenta que en las PCs de Facultad el acceso a Internet se realiza a través de un servidor proxy. Esto puede generar diferencias con pruebas en otras redes.
- Para facilitar los análisis con la herramienta Wireshark, pruebe de aplicar filtros a la captura de paquetes que realiza.

[1]Deberá entregar el archivo de dicha descarga como parte de esta entrega.