



Trabajo Práctico 1

Wiretapping

21 de abril de 2016

Teoría de las comunicaciones

Integrante	LU	Correo electrónico
Barbeito, Nicolás	147/10	barbeiton@yahoo.com.ar
Interlandi, Daniel	773/00	danielinterlandi@yahoo.com.ar
Ladelfa, Hernán Nahuel	318/04	nahueladelfa@gmail.com

Instancia	Docente	Nota
Primera entrega		
Segunda entrega		



Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (++54 +11) 4576-3300

<http://www.exactas.uba.ar>

Índice

1. Resumen	2
2. Introducción	2
3. Herramienta desarrollada	3
4. Resultados y análisis	4
4.1. Red Doméstica	4
4.1.1. Paquetes capturados e información	4
4.1.2. Nodos de la red	5
4.1.3. Análisis de la de entropía	5
4.2. Red Laboral	6
4.2.1. Paquetes capturados e información	6
4.2.2. Análisis de la de entropía	6
4.3. Red de Centro Comercial	7
4.3.1. Paquetes capturados e información	7
4.3.2. Nodos de la red	8
4.3.3. Análisis de la de entropía	8
5. Conclusiones	9

1. Resumen

El objetivo de este trabajo es utilizar técnicas provistas por la teoría de la información para distinguir diversos aspectos de la red de manera analítica. Además, sugerimos el uso de dos herramientas modernas de manipulación y análisis de paquetes frecuentemente usadas en el dominio de las redes de computadoras: Wireshark y Scapy.

2. Introducción

Se modelaron las redes analizadas como dos fuentes, S y S_1 . La primera tiene como objetivo examinar los protocolos presentes y su importancia en la misma, se toma

$$S = \{s_1 \dots s_n\}$$

donde s_i es $p_i.type$, con p_i el i-ésimo paquete capturado.

El segundo modelo es

$$S_1 = \{s_{1,1} \dots s_{1,n}\}$$

donde $s_{1,i}$ es $p_i.pdst$ (dirección IP destino), con p_i el i-ésimo paquete del protocolo ARP capturado. Este es utilizado para analizar la importancia de los nodos dentro de la red.

Usando estos modelos y los conceptos de información y entropía podemos detectar símbolos destacados en las fuentes (nodos o protocolos según la fuente). La información que otorga un símbolo s_i en una fuente se define como

$$I(s_i) = \log(1/P(s_i))$$

y nos ayuda a detectar que tan frecuente es la emisión de ese símbolo por la fuente. La entropía de una fuente $S = \{s_1 \dots s_n\}$ se define como

$$\sum_S P(s_i)I(s_i) \forall s_i \in S$$

este número da la información media emitida por la fuente. Indica que tan desordenada es la aparición de los símbolos.

Un concepto fundamental es el del protocolo ARP. Este sirve para relacionar direcciones de nivel de red (IP) con direcciones de nivel de enlace (MAC). Cuando un equipo desea mandar un paquete a una dirección IP dada, necesita ubicar el mismo a nivel de enlace, entonces envía un ARP-request mediante broadcast requiriendo una respuesta del poseedor de la dirección IP, luego, si existe, el nodo que tenga esa IP responderá usando un paquete ARP reply a quién realizó la consulta con su dirección MAC. De esta manera, interceptando paquetes ARP se puede detectar qué nodos activos hay en una red.

3. Herramienta desarrollada

Para realizar las capturas de las redes, se desarrolló un programa en lenguaje Python integrado con la herramienta Scapy. Este programa captura el tráfico de la red en modo promiscuo. La información de los paquetes leídos es almacenada en diferentes archivos para su posterior análisis.

En estos archivos se almacenan datos como, la cantidad de paquetes, la información, la probabilidad y la entropía. Estos datos están relacionados con las dos fuentes elegidas: por tipo de protocolo para los paquetes Ethernet y por IP para la captura de paquetes de tipo ARP.

Además, parte de estos datos también se almacenan en un archivo con formato JSON. Este archivo es utilizado por otro programa también desarrollado en Python para generar gráficos y así poder analizar la información capturada de la red. Este programa genera gráficos de tipo torta, histogramas y grafos de la red según los paquetes ARP enviados o recibidos.

Estos programas se pueden ejecutar desde una terminal con sistema operativo Linux u OS X, con lenguaje Python y las librerías Scapy y Matplotlib.

Para la ejecución de la herramienta de captura de paquetes de la red, se debe ejecutar el siguiente comando:

```
$ sudo ./sniffer.py <file_prefix> <timeout>
```

Para la ejecución del programa que genera los gráficos, se debe ejecutar:

```
$ sudo ./plot.py <file_prefix>
```

Donde:

- file_prefix: Es el prefijo que se le agregará al nombre del archivo JSON generado por el programa sniffer.py y el prefijo con el cual plot.py lo ubicará. Además es el prefijo que plot.py le colocará a los nombres de los archivos de los gráficos generados.
- timeout: Es el tiempo durante el cual el programa se encontrará capturando tráfico de la red.

4. Resultados y análisis

4.1. Red Doméstica

Para la primera captura, se eligió la red doméstica de uno de los integrantes del grupo. Los dispositivos conectados a la red en este caso fueron, 3 computadoras, 2 teléfonos celulares, un televisor SmartTV y un Apple TV. Todos estos conectados al modem del proveedor de internet. La captura se realizó desde la computadora con dirección IP 192.168.0.30 mediante conexión WIFI y la misma duró aproximadamente 30 minutos.

4.1.1. Paquetes capturados e información

A continuación analizaremos la relación entre la cantidad de paquetes y la información que provee cada tipo de protocolo de la fuente S .

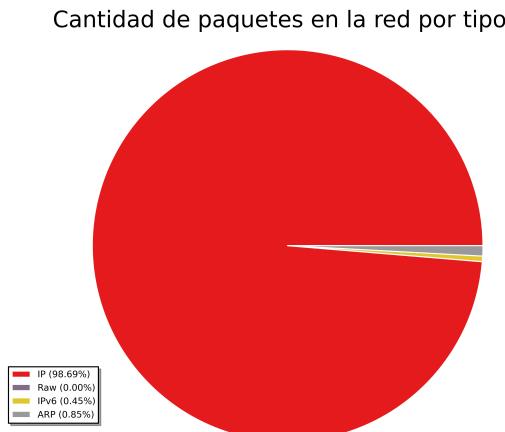


Figura 1: Fuente S

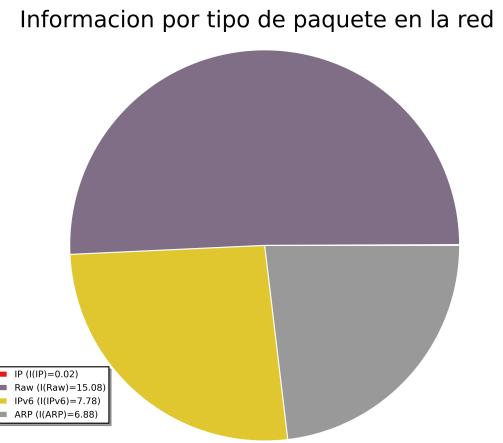


Figura 2: Fuente S

En los gráficos Figura 1. y Figura 2. se puede observar que el protocolo que presenta mayor frecuencia es el IPv4 con un porcentaje muy superior al resto y que, por el contrario, aporta muy poca información. Por lo tanto, en este caso, el símbolo distinguido en la fuente S sería el que representa al tipo de paquete IPv4.

Podemos observar que, para este caso, la incidencia de los paquetes ARP en la red es baja en comparación al resto de los protocolos.

Ahora veremos la relación entre la cantidad de paquetes y la información que proporciona cada nodo en la red para la fuente S_1 .

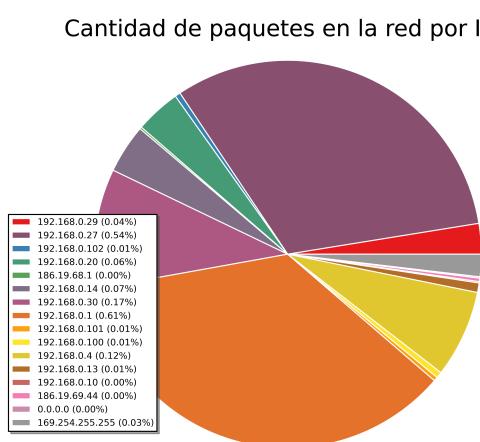


Figura 3: Fuente S_1

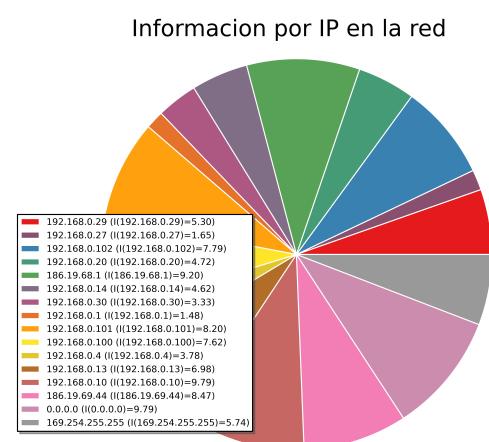


Figura 4: Fuente S_1

En los datos de los gráficos Figura 3. y Figura 4. se toma como fuente a las direcciones IP de la red. Se puede observar que los nodos distinguidos de la red para la fuente S_1 son los que poseen las direcciones IP 192.168.0.1 y 192.168.0.27, ya que son los que presentan mayor frecuencia y, por lo tanto, los que menos información aportan.

4.1.2. Nodos de la red

A continuación, en la Figura 5. se pueden observar los diferentes nodos de la red asociados a sus direcciones IP. Los ejes que conectan a un par de nodos representan que entre ellos hubo algún envío de paquetes ARP. El tamaño de cada nodo es proporcional a la cantidad de paquetes que el mismo envió y recibió.

Topografia de la red segun paquetes ARP enviados

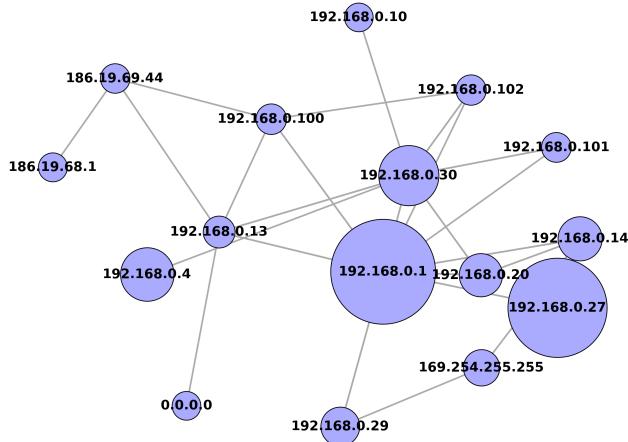


Figura 5

Al ser una red pequeña podemos confirmar claramente cuales son los nodos distinguidos: El que posee la dirección IP 192.168.0.1 que corresponde al modem y el host 192.168.0.27 correspondiente al SmartTV, que en el momento de la captura de los paquetes, el mismo se encontraba realizando actualizaciones.

4.1.3. Análisis de la de entropía

A continuación analizaremos histogramas con cortes en los valores de entropía, tanto para las IP de la red como para los tipos de protocolo.

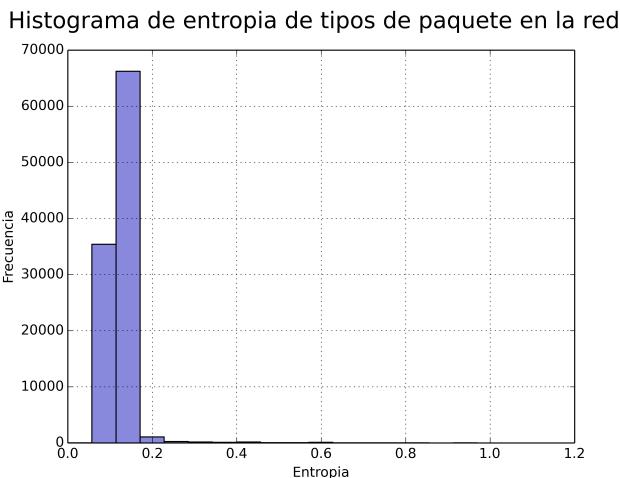


Figura 6: Fuente S

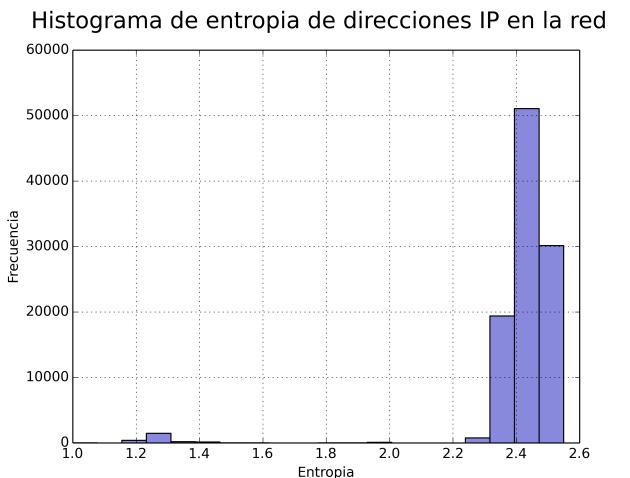


Figura 7: Fuente S₁

Podemos observar que para el caso de la captura de paquetes de la fuente S_1 se presenta una entropía media mayor que en la fuente S . Esto se debe a la impredecibilidad de los símbolos en el caso de las direcciones IP en comparación con la de los tipos de protocolo, donde la mayoría de los protocolos de los paquetes fueron de tipo IPv4.

4.2. Red Laboral

Esta captura se realizó en la red laboral de uno de los integrantes del grupo. En este caso se puede notar la gran cantidad de nodos y tráfico de paquetes. El gráfico de topología era tan extenso que decidimos no publicarlo.

4.2.1. Paquetes capturados e información

Analizaremos la relación entre la cantidad de paquetes y la información que provee cada tipo de protocolo de la fuente S .

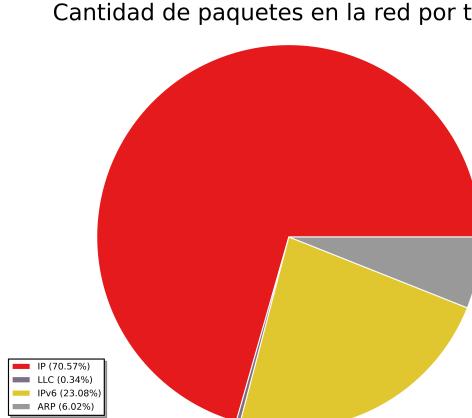


Figura 8: Fuente S

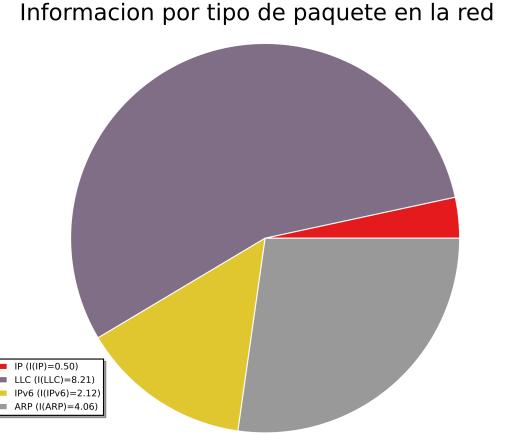


Figura 9: Fuente S

En los gráficos Figura 8. y Figura 9 se puede observar que el protocolo IPv4 es el más frecuente. Pero también se observa gran cantidad de paquetes IPv6.

Podemos ver que la cantidad de paquetes ARP, en comparación con IPv4 e IPv6, es baja. Sin embargo, en este caso, la proporción de paquetes ARP es mayor que en el caso de la red doméstica.

Veremos la relación entre la cantidad de paquetes y la información que proporciona cada nodo en la red para la fuente S_1 .

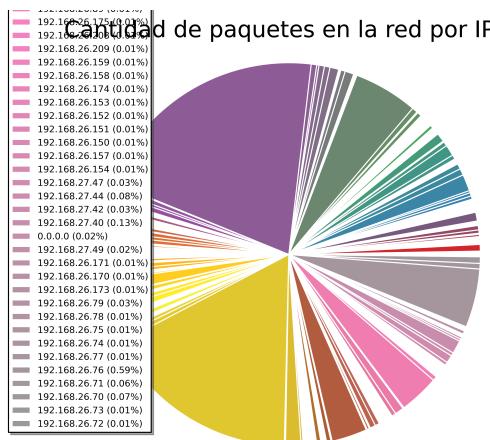


Figura 10: Fuente S_1

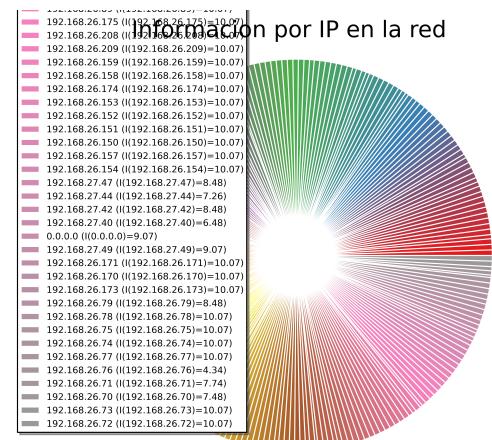


Figura 11: Fuente S_1

Si bien se trata de una red grande, se puede ver en los gráficos Figura 10. y Figura 11. que hay dos nodos que distinguen del resto: El que tiene dirección IP 192.168.26.43 con el 0.2% y el de IP 192.168.26.1 con el 0.17%. Se presume que esos nodos deben ser routers.

4.2.2. Análisis de la entropía

Vemos histogramas con cortes en los valores de entropía, tanto para las fuentes S y S_1

Histograma de entropía de direcciones IP en la red

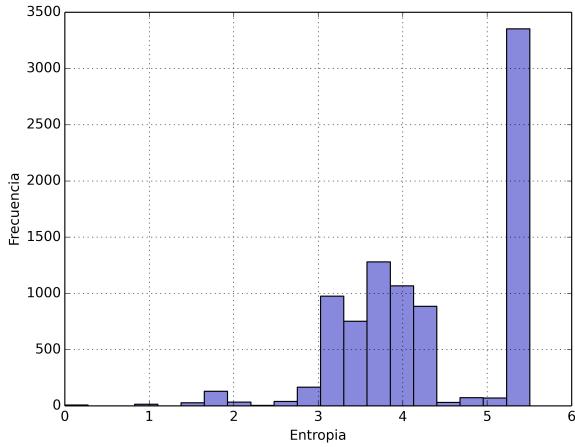


Figura 12: Fuente *S*

Histograma de entropía de direcciones IP en la red

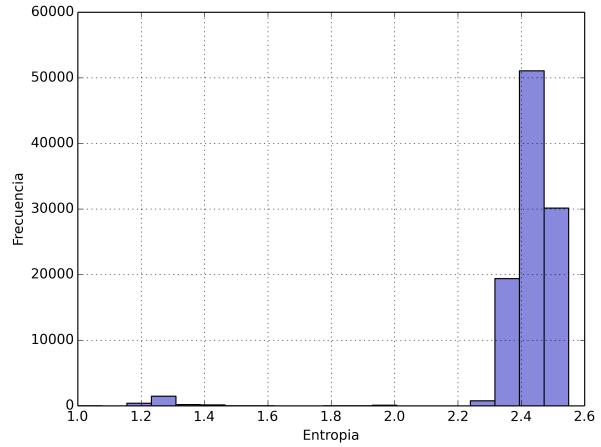


Figura 13: Fuente *S*₁

4.3. Red de Centro Comercial

En este caso la captura se llevó a cabo en una red no controlada, en el shopping Galerías Pacífico, durante aproximadamente 10 minutos. Debido a que la red no está controlada por nosotros, no conocemos la naturaleza de los equipos que intercambian información en la misma, pero conjeturamos que la mayoría son teléfonos celulares.

4.3.1. Paquetes capturados e información

Los resultados del experimento para determinar protocolos importantes para la fuente *S* se resumen en los siguientes gráficos.

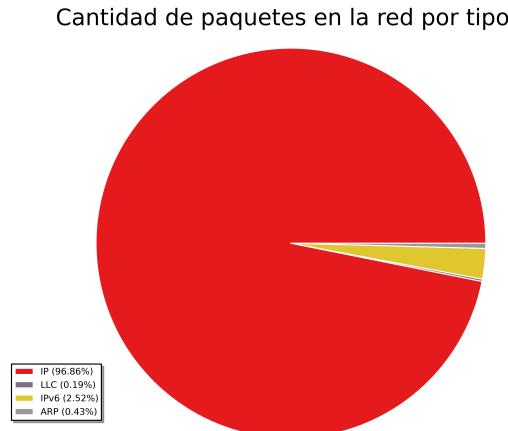


Figura 14: Fuente *S*

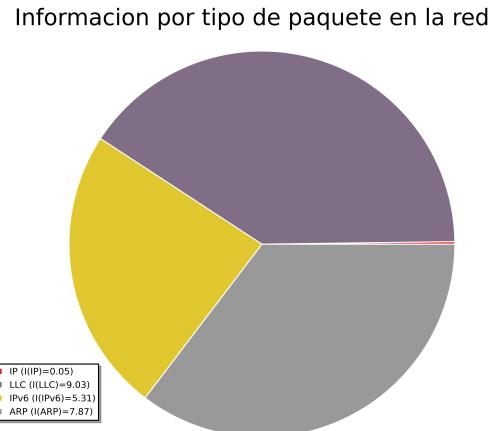


Figura 15: Fuente *S*

En los gráficos Figura 14. y Figura 15. se puede observar que el protocolo IPV4 es el más frecuente y, por ende, el que brinda menos información.

Mostramos ahora la frecuencia e información de cada IP en la red para la fuente *S*₁.

En los gráficos Figura 16. y Figura 17. podemos observar que las direcciones IP distinguidas son 192.104.200.5, 190.104.200.1 y 190.104.200.188.

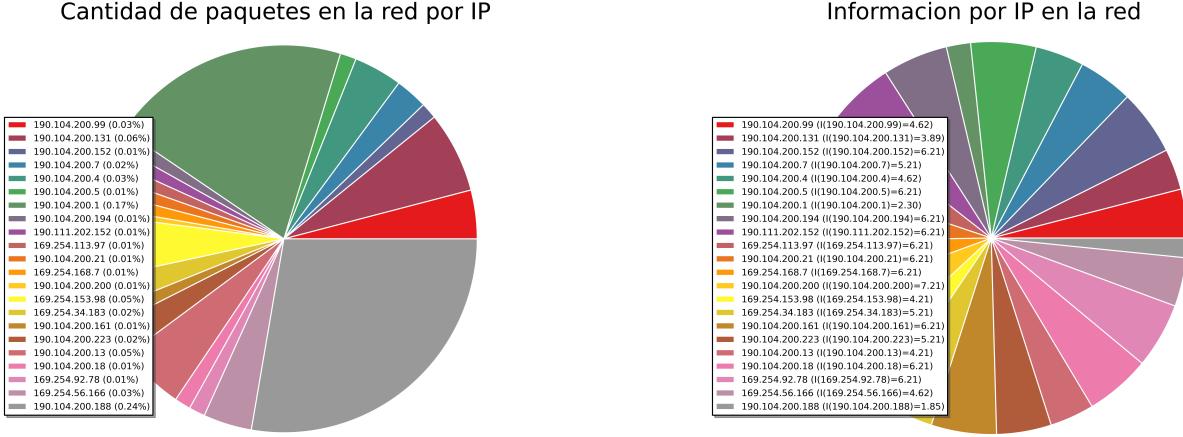


Figura 16: Fuente S_1

Figura 17: Fuente S_1

4.3.2. Nodos de la red

En la Figura 18 se muestra un grafo con los diferentes nodos en la red identificados por su dirección IP.

Topografía de la red segun paquetes ARP enviados

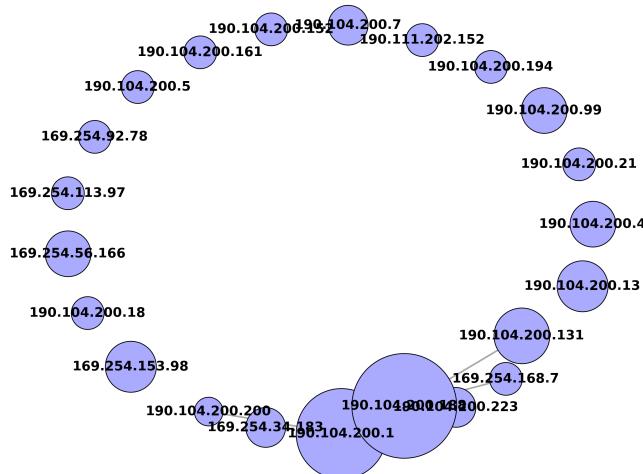


Figura 18

4.3.3. Análisis de la entropía

A continuación analizaremos histogramas con cortes en los valores de entropía, tanto para las IP de la red como para los tipos de protocolo.

Al igual que en las otras redes, podemos ver que la entropía en la fuente S es menor que en la fuente S_1 , debido a que en la primera, los símbolos son más predecibles.

Histograma de entropía de direcciones IP en la red

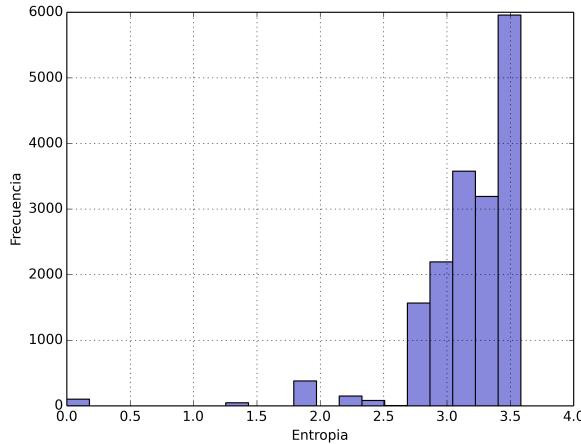


Figura 19: Fuente S

Histograma de entropía de tipos de paquete en la red

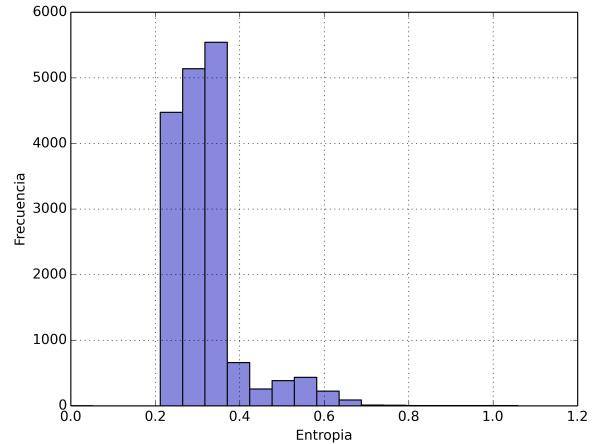


Figura 20: Fuente S_1

5. Conclusiones

En este experimento pudimos apreciar una aplicación de la teoría de la información para poder descubrir partes de una red cuyo impacto es más destacado en la misma, en particular, nodos y protocolos con mayor importancia. En la primera parte del trabajo práctico, al buscar protocolos distinguidos, se descubrió que, en general, el protocolo más frecuente es el IPv4, lo cual es razonable. Además se notó que en algunas redes el protocolo IPv6, también, es bastante frecuente y que el protocolo ARP siempre se encontró presente en ellas. Esto último es previsible, por el funcionamiento de las redes IP en LAN.

En la segunda parte del trabajo se investigó la existencia de nodos distinguidos (o símbolos distinguidos en el contexto de la fuente S_1). Se detectó que en cada una de las redes el router fue uno de los nodos distinguidos. También se observó que en las redes no controladas la entropía de la fuente S_1 es menor que en aquellas que sí son controladas, esto se puede deber a que la entropía mide la previsibilidad de la fuente, por lo que en redes no controladas esta será mayor.