



Trabajo Práctico 1

Wiretapping

20 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. Resultados y análisis	3
3.1. Herramienta	3
3.2. Red Doméstica	4
3.2.1. Paquetes capturados e información	4
3.2.2. Histogramas (de IPs y protocolos)	4
3.3. Red Laboral	5
3.3.1. Histogramas (de IPs y protocolos)	8
3.3.2. Paquetes capturados e información	8
3.4. Red de Galerías Pacífico	9
3.4.1. Paquetes capturados e información	12
3.4.2. Histogramas (de IPs y protocolos)	13
4. Conclusiones	14

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. Resultados y análisis

3.1. Herramienta

Para realizar las capturas de las redes, se desarrollo un script en el lenguaje Python integrado con la herramienta Scapy. Este script escucha pasivamente la red capturando los paquetes que son almacenados en distintos archivos para su posterior análisis.

En los archivos se guarda la información, probabilidad y entropía de las 2 fuentes elegidas. Por tipo de protocolo para los paquetes Ethernet y por IP para la captura de paquetes ARP.

Un 3 archivo de formato json, contiene la información de manera organizada para poder ser leído por otro script desarrollado utilizado para generar distintos gráficos.

Este segundo script genera gráficos como histogramas, de torta y grafos según paquetes ARP enviados.

Para la ejecución de la herramienta, debemos ejecutar los siguiente comandos desde una terminal en el sistema operativo Linux.

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

Para la ejecución del graficador:

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

3.2. 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 duró aproximadamente 30 minutos. En Figura 1. se pueden observar los diferentes nodos de la red asociados a su dirección IP. Los ejes que conectan a un par de nodos representan que ellos hubo algún envío de paquetes. El tamaño de cada nodo es proporcional a la cantidad de paquetes el mismo envió y recibió.

Al ser una red pequeña podemos distinguir claramente a los nodos distinguidos. El que posee la dirección IP 192.168.0.1 que corresponde al modem y el nodo 192.168.0.27 correspondiente al SmartTV, que en el momento de la captura de los paquetes, el mismo se encontraba realizando una actualización.

Topografía de la red segun paquetes ARP enviados

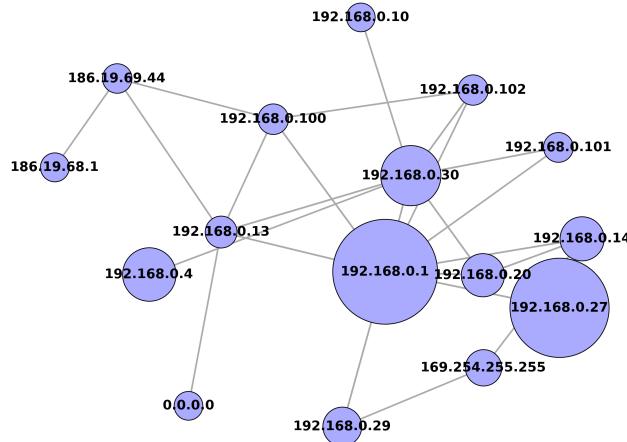


Figura 1

3.2.1. Paquetes capturados e información

Los gráficos de torta nos permiten ver la relación entre la cantidad de paquetes y la información que provee cada nodo en la red.

En los gráficos 2. y 3. se toma como fuente a las direcciones IP de la red. Se puede observar que los nodos distinguidos mencionados anteriormente son los mas frecuentes y por lo tanto los que menos información presentan. A continuación analizaremos la relación entre la cantidad de paquetes y la información que provee cada tipo de paquete.

En los gráficos 4. y 5. la fuente es la indicada en la cátedra. Se observa que el protocolo que presenta mayor frecuencia es el IP con un porcentaje muy superior al resto y que por el contrario, aporta muy poca información. En este caso el símbolo distinguido en esta fuente sería el que representa al tipo de paquete IP.

3.2.2. Histogramas (de IPs y protocolos)

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

Podemos observar que para el caso de la captura de paquetes de la fuente de IPs se presenta una entropía media mayor que en la fuente por tipo protocolo. Esto se debe a la impredecibilidad de los símbolos en el caso de las IP en comparación con la de los protocolos, donde la mayoría de los paquetes eran de tipo IP.

Cantidad de paquetes en la red por IP

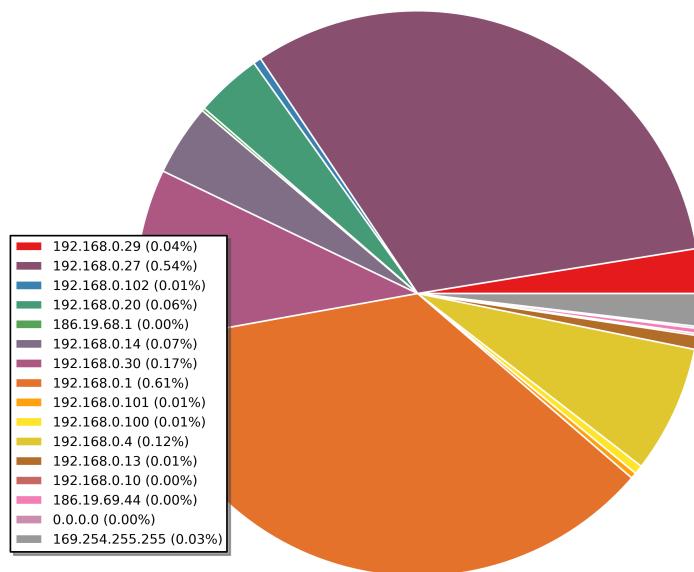


Figura 2

Informacion por IP en la red

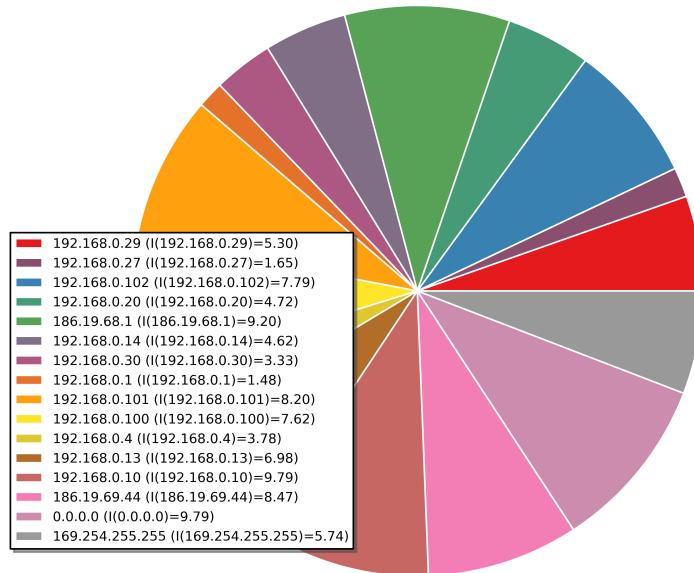


Figura 3

3.3. Red Laboral

Con las capturas de esta red, se puede notar la gran cantidad de nodos y trafico de paquetes. El gráfico de topología era tan extenso que decidimos no ponerlo.

Cantidad de paquetes en la red por tipo

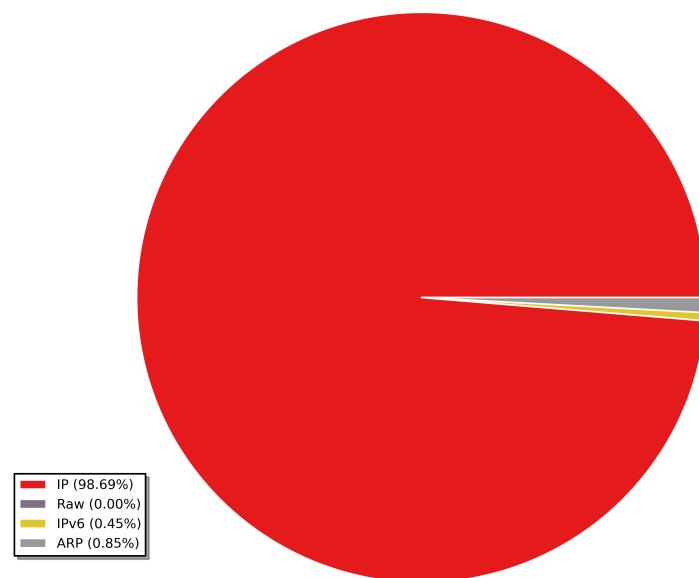


Figura 4

Informacion por tipo de paquete en la red

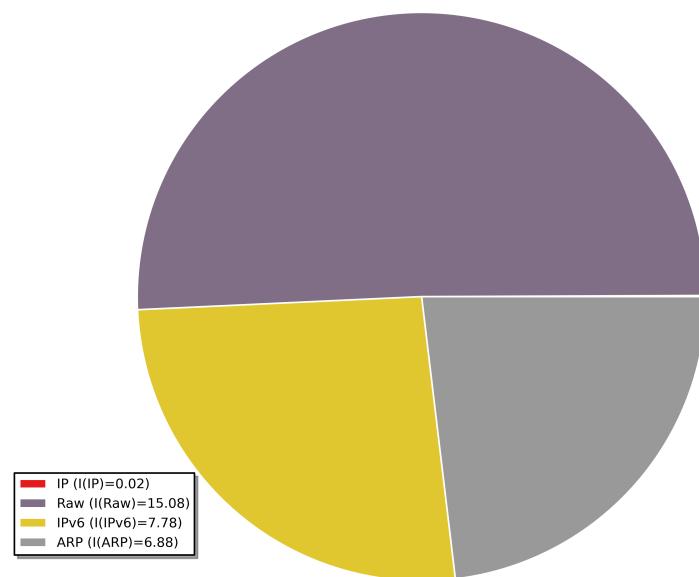


Figura 5

Histograma de entropia de direcciones IP en la red

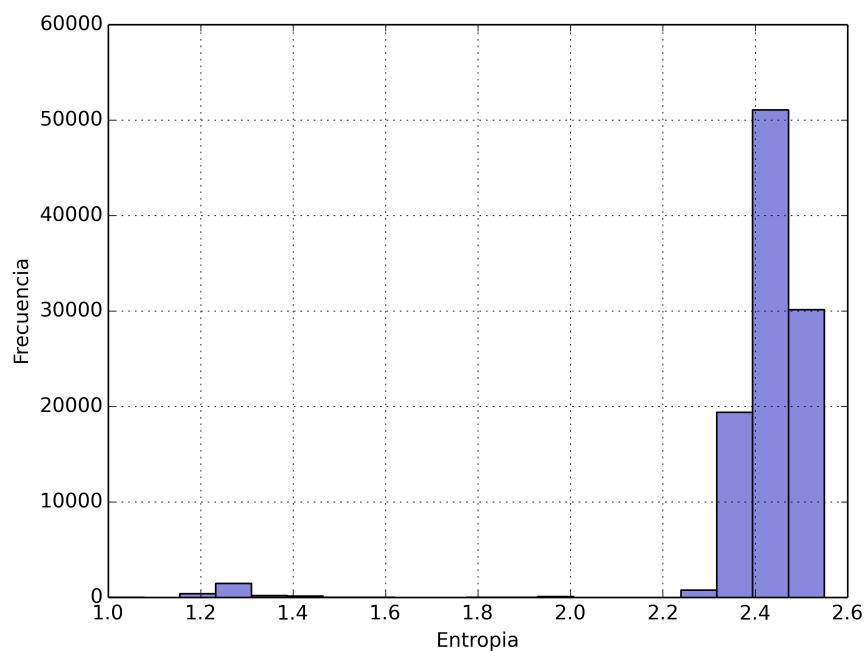


Figura 6: Mi Figura

Histograma de entropia de tipos de paquete en la red

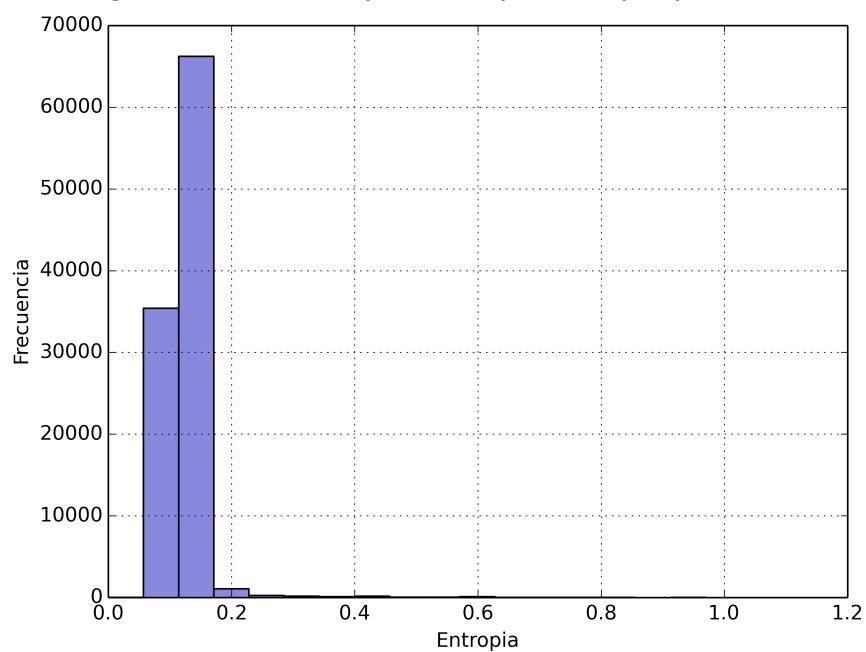


Figura 7: Mi Figura

3.3.1. Histogramas (de IPs y protocolos)

Histograma de entropía de direcciones IP en la red

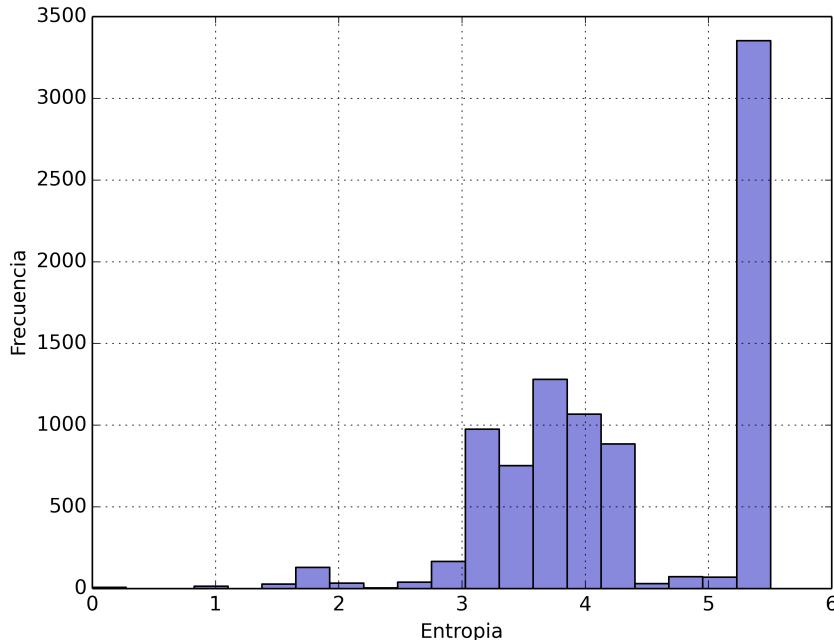


Figura 8: Mi Figura

Histograma de entropía de tipos de paquete en la red

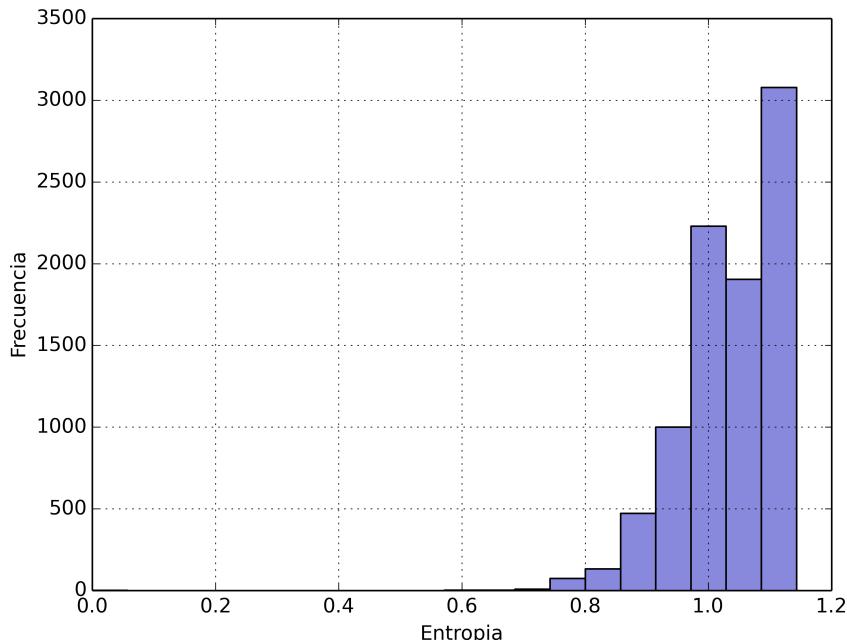


Figura 9: Mi Figura

3.3.2. Paquetes capturados e información

Si bien es una red grande, se puede ver en los gráficos que hay 2 nodos que distinguen del resto 192.168.26.43 con el 0.2% y el 192.168.26.1 con el 0.17%. Se presume que esos nodos tienen que ser routers [10](#) [11](#).

En los gráficos de protocolos se vuelve a repetir que el protocolo IP es el mas frecuente. Pero también se observa gran cantidad de paquetes IPv6 [12](#) [13](#).

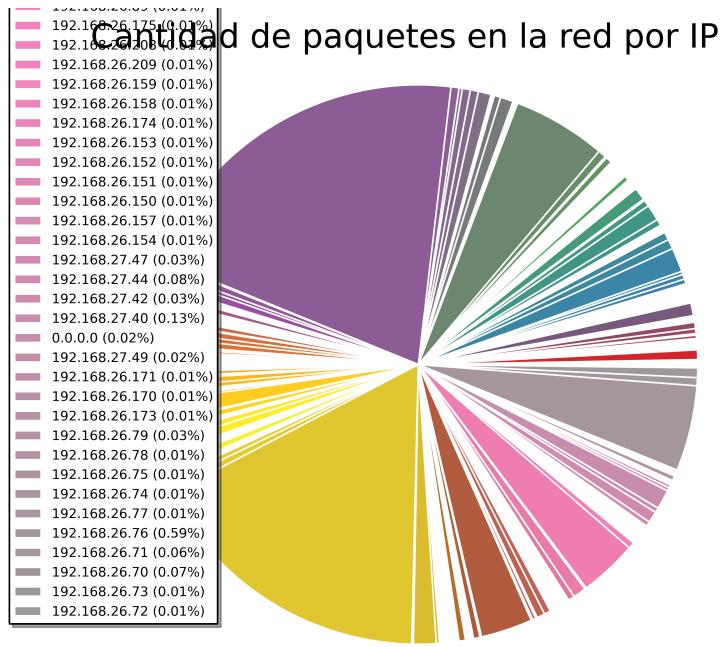


Figura 10: Mi Figura

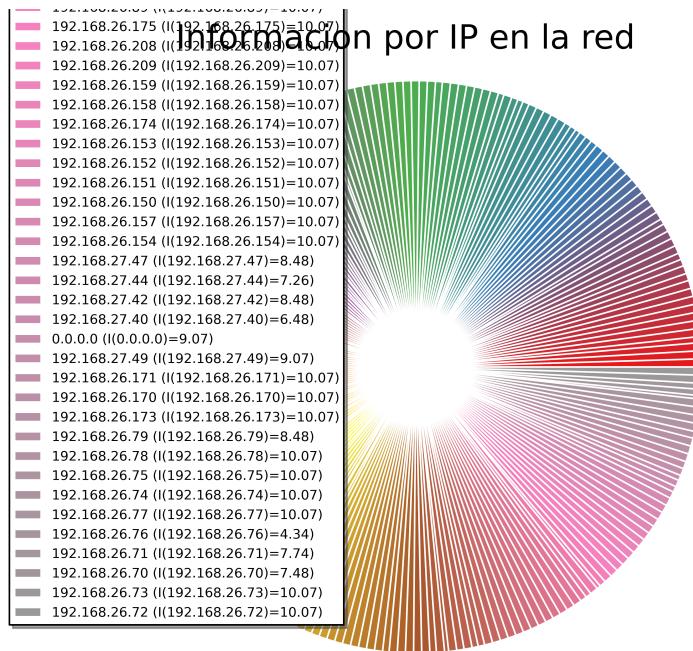


Figura 11: Mi Figura

3.4. Red de Galerías Pacífico

En este caso la captura se llevó a cabo en una red no controlada, en Galerías Pacífico, durante aproximadamente 10 minutos. Debido a que la red no está controlada por nosotros, no sabemos la naturaleza de los equipos que intercambian información en la misma, pero conjeturamos que la mayoría son celulares. Este [14](#) es un grafo que muestra los distintos nodos en la red junto identificados por su dirección IP.

Cantidad de paquetes en la red por tipo

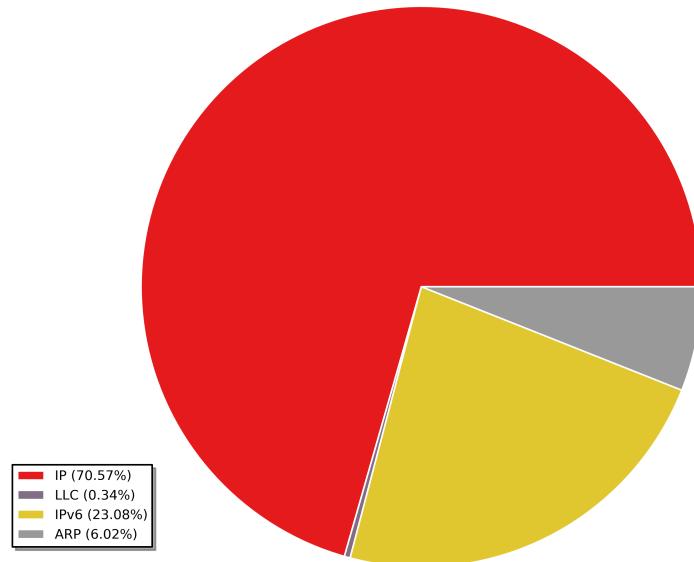


Figura 12: Mi Figura

Informacion por tipo de paquete en la red

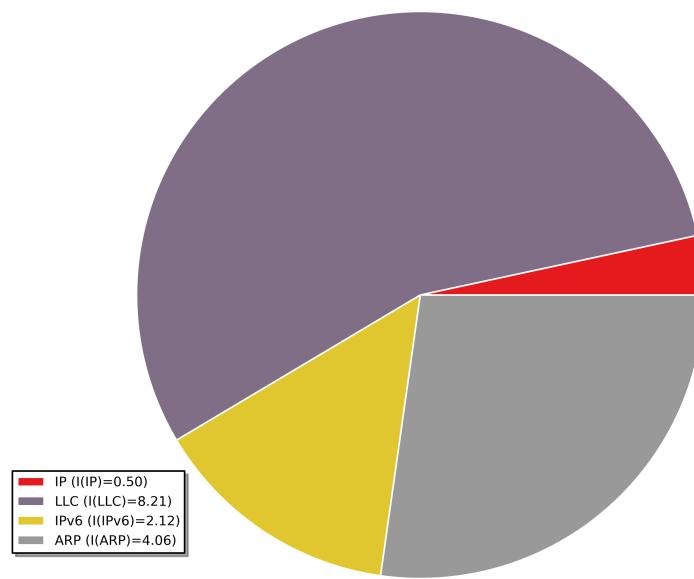


Figura 13: Mi Figura

Topografia de la red segun paquetes ARP enviados

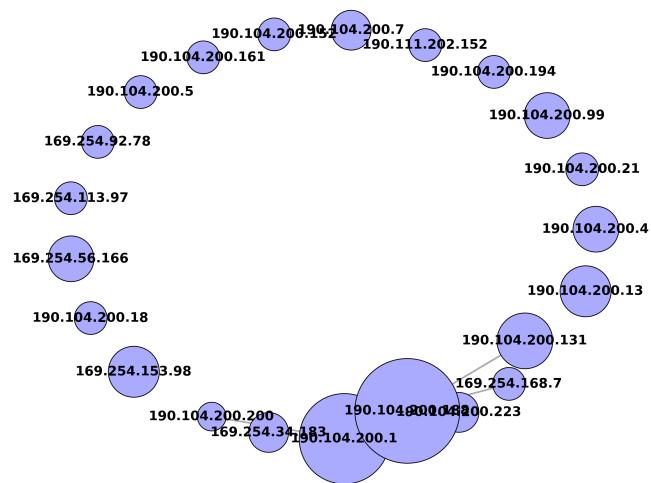


Figura 14

3.4.1. Paquetes capturados e información

Mostramos ahora la frecuencia e información de cada IP en la red en [15](#) y [16](#).

Cantidad de paquetes en la red por IP

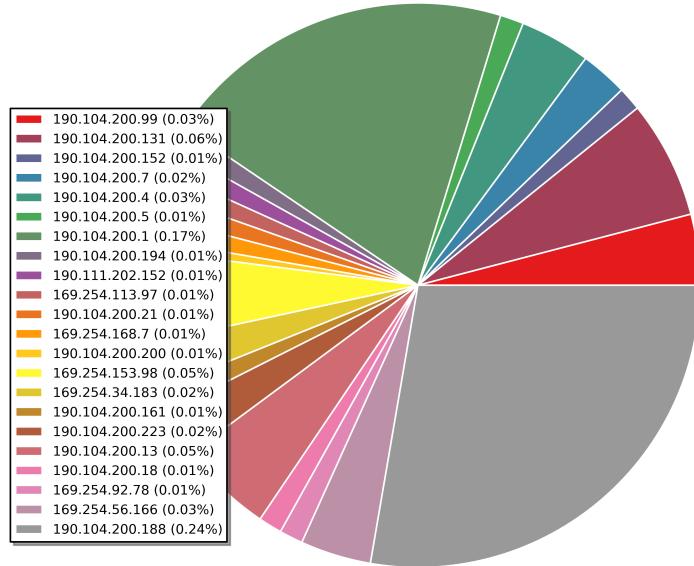


Figura 15

Informacion por IP en la red

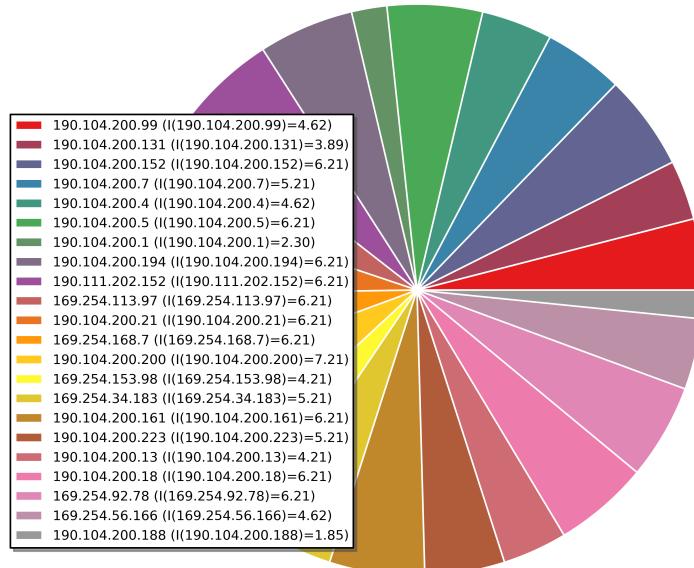


Figura 16

Podemos ver que las IP's distinguidas son 192.104.200.5, 190.104.200.1 y 190.104.200.188.

Los resultados del experimento para determinar protocolos importantes se resumen en los gráficos [17](#) y [18](#). Se puede notar que el protocolo IPV4 es el más frecuente (y por ende el que brinda más información).

Cantidad de paquetes en la red por tipo

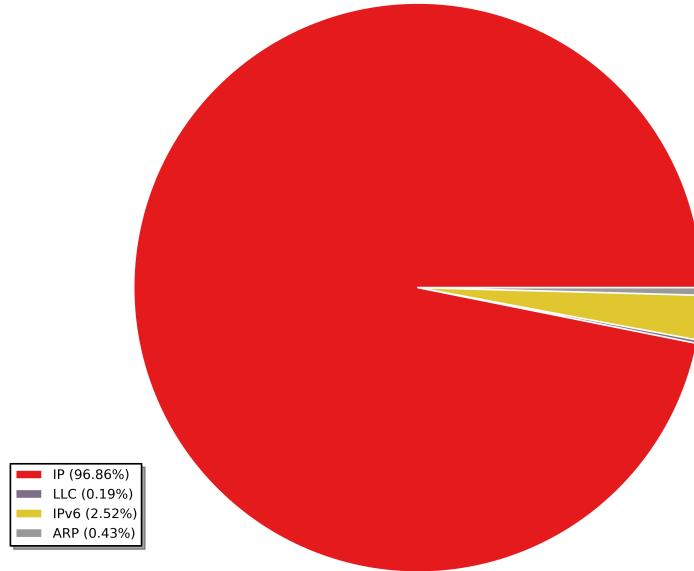


Figura 17

Informacion por tipo de paquete en la red

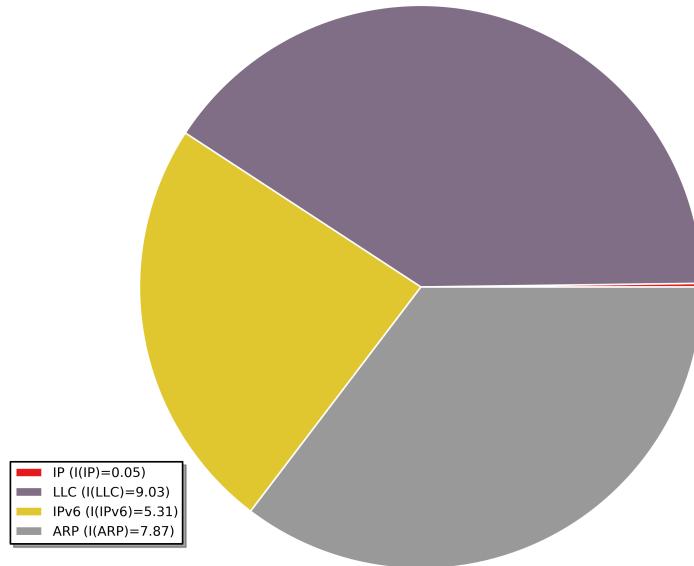


Figura 18

3.4.2. Histogramas (de IPs y protocolos)

En esta sección usamos el concepto de entropía para deducir características de la red analizada. Al igual que en las otras redes, podemos ver en [20](#) y en [19](#) que la entropía en la fuente S es menor que en la fuente S_1 , debido a que en la primera, el protocolo IPV4 es mucho más frecuente que los demás.

Histograma de entropía de direcciones IP en la red

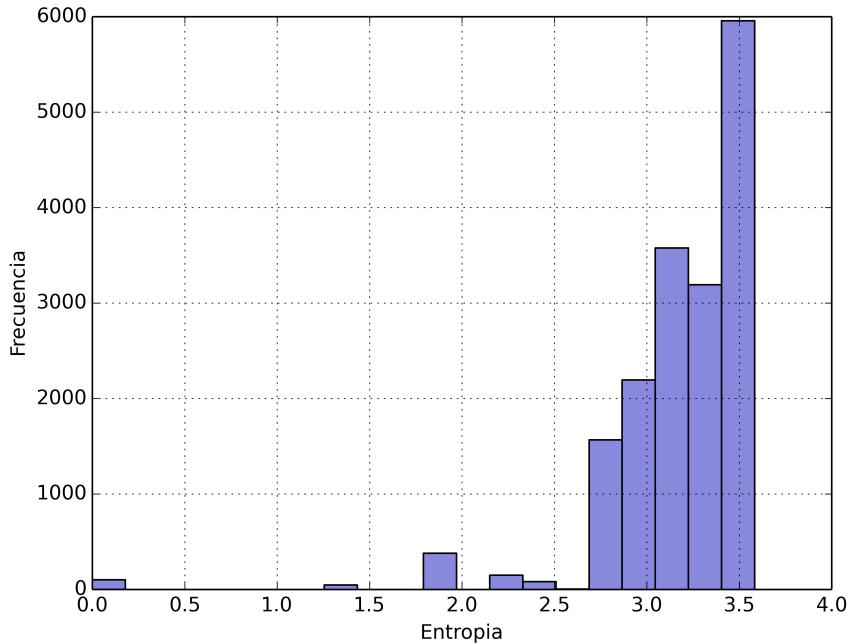


Figura 19

Histograma de entropía de tipos de paquete en la red

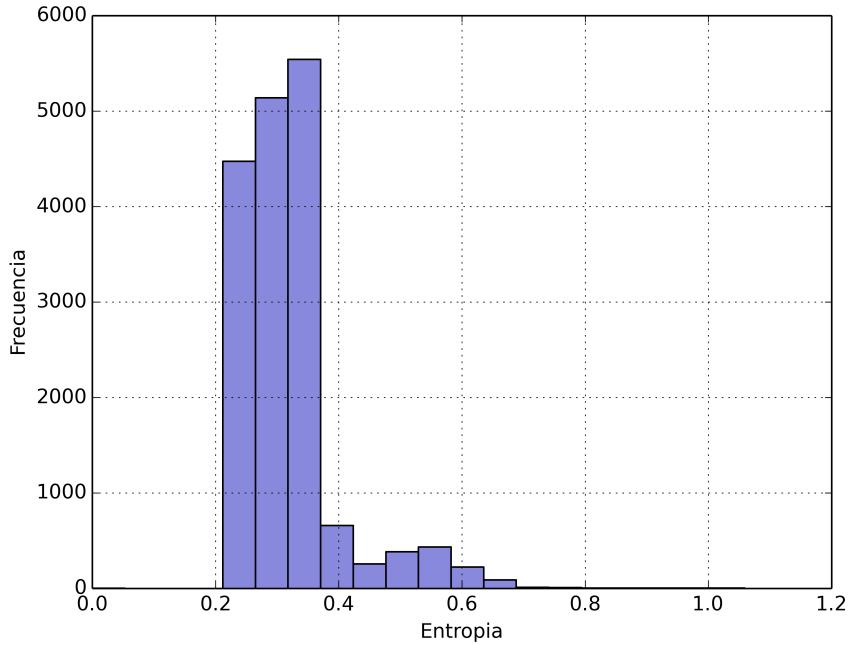


Figura 20

4. 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 primer parte del trabajo, al buscar protocolos diferenciados, se descubrió que el protocolo más usado es el de IPV4, lo que es razonable. Además se notó que en algunas redes el protocolo IPV6 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 símbolo distinguido 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 éstas será mayor.