DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 1

Wiretapping

15 de julio de 2016 — Teoría de las comunicaciones

Integrante	LU	Correo electrónico
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. Resúmen

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 de información, S y S_1. La primera tiene como objetivo examinar los protocolos presentes y su relevancia en la misma, se toma

$$S = \{s_1, ..., s_n\}$$

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

El segundo modelo es

$$S_1 = \{s_{1_1}, ..., s_{1_n}\}$$

donde s_{1i} es $p_i.psrc$ (dirección IP origen) ó $p_i.pdst$ (dirección IP destino), con p_i el i-ésimo paquete del protocolo ARP capturado.

La información que otorga un símbolo s_i en una fuente se define como

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

Donde $P(s_i)$ es la probabilidad del símbolo s_i. Esto es utilizado para analizar la información que aporta cada nodo dentro de la red.
La entropía de una fuente $S = \{s_1, ..., s_n\}$ 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 qué tan desordenada es la aparición de los símbolos.

Utilizando estos modelos y los conceptos de información y entropía podremos detectar símbolos distinguidos en las fuentes. Llamamos símbolos distinguidos a es aquellos para los que su información dista mucho de la entropía de la fuente.

Otro 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 envia 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.

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_preffix> <timeout>
```

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

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

Donde:

- file_preffix: 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

Realizamos experimentos en diversas redes con el objetivo de analizar ciertos aspectos en base a la información que aportaron los paquetes modelados como diferentes símbolos de una fuente.
Se realizaron capturas de paquetes, con la herramienta desarrollada, en distintas redes. Con los datos obtenidos se generaron diversos gráficos para realizar un análisis del comportamiento de la red en cada experimento.
Buscaremos encontrar e interpretar a los protocolos y nodos distinguidos utilizando la información que aportaron en las capturas. Además analizaremos la incidencia que presentaron los paquetes de tipo ARP en la red.

Para cada experimento se generaron grafos de la topología de la red capturada y gráficos que muestran la relación entre la cantidad de información de los símbolos y la entropía.

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 Cisco DPC2420 provisto por el 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.

A continuación 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 él mismo envió y/o recibió. Esto nos sirve para darnos una idea de la topología de la red.

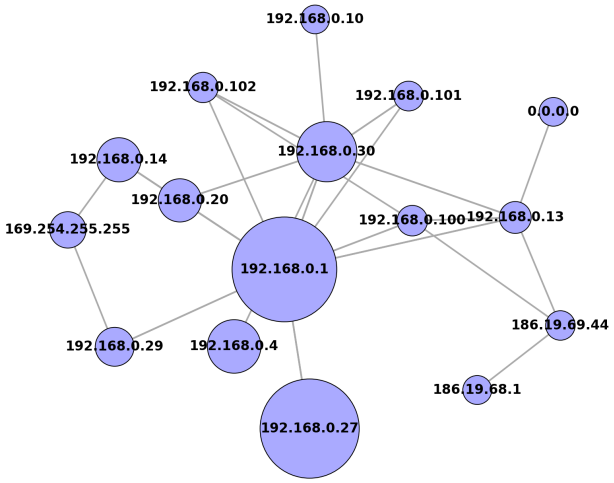

Figura 1

Al ser una red pequeña podemos observar cuáles podrían ser los nodos distinguidos. Uno el que posee la dirección IP 192.168.0.1 que corresponde al modem/router del proveedor de internet. Y el host 192.168.0.27 correspondiente al SmartTV, que según lo que investigamos acerca de su modelo, realiza constantemente búsquedas de nuevos dispositivos conectados a la red.

A continuación veremos un gráfico que presenta la cantidad de información de cada nodo en la red. Mediante una recta horizontal se representa el valor de la entropía.

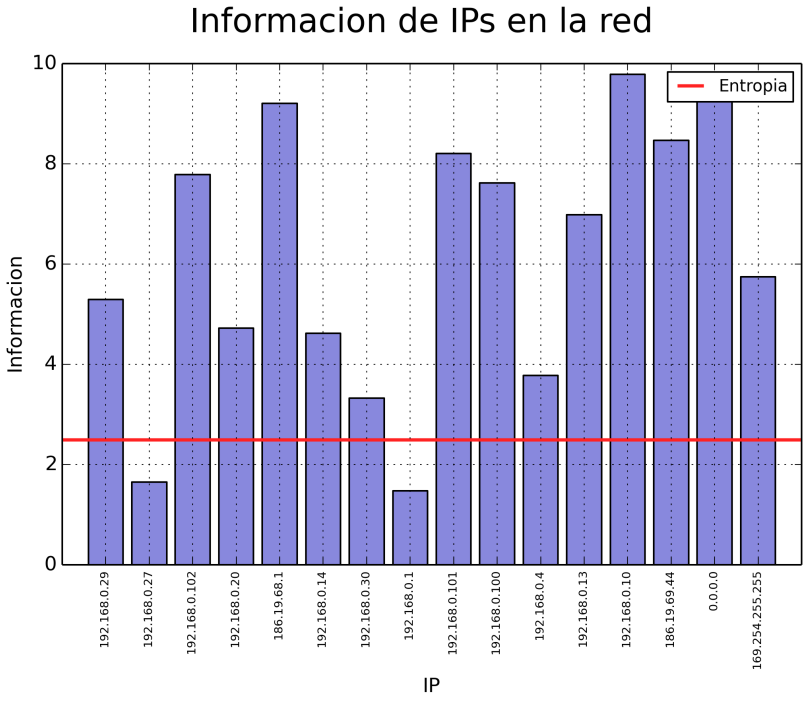

Figura 2

Podemos observar que los nodos mencionados anteriormente, los que tienen dirección IP 192.168.0.1 y 192.168.0.27, son nodos distinguidos ya que son los únicos para los que su información se encuentra por debajo de la entropía. Es decir que presentaron una alta cantidad de paquetes de tipo ARP en la captura. Por el contrario se pueden ver otros nodos que aportaron mucha información en comparación a la entropía. Este podría ser el caso de algún dispositivo que haya tenido poca participación en el descubrimiento de interfaces mediante el protocolo ARP.

A continuación veremos un gráfico mostrando la cantidad de información capturada de cada tipo de protocolo en comparación con el valor de la entropía.

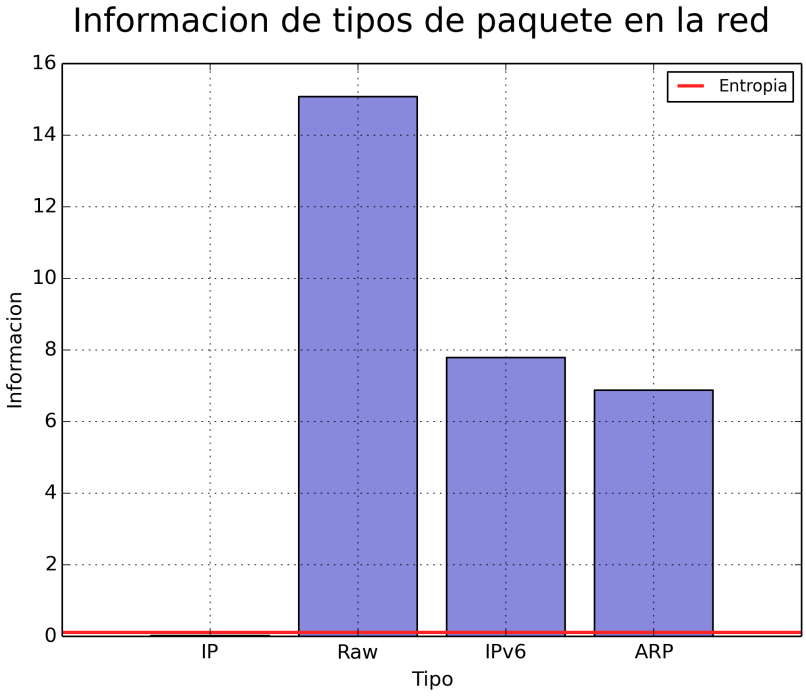

Figura 3

Podemos observar que, para este caso, la incidencia de los paquetes ARP en la red es baja en comparación al protocolo IP. Lo mismo ocurre para los protocolos IPv6 y RAW. Es decir que el protocolo IP aporta menor cantidad

de información en comparación con los otros tres protocolos. De todas formas, El protocolo ARP tuvo mayor frecuencia, y aportó menor información, que los protocolos IPv6 y RAW. Esto puede darse dado que estos últimos tienen menor uso que ARP.

4.2. Red Laboral

Esta captura se realizó en la red laboral de uno de los integrantes del grupo. Al ser una red con una gran cantidad de nodos y tráfico de paquetes, se optó por hacer las capturas al final de la jornada laboral donde el tráfico es menor. En otro momento los gráficos no resultaban claros para su análisis. La captura se realizó durante aproximadamente 10 minutos.

Vemos a continuación el gráfico que muestra la topología de la red según el intercambio de paquetes ARP que hubo en la captura.

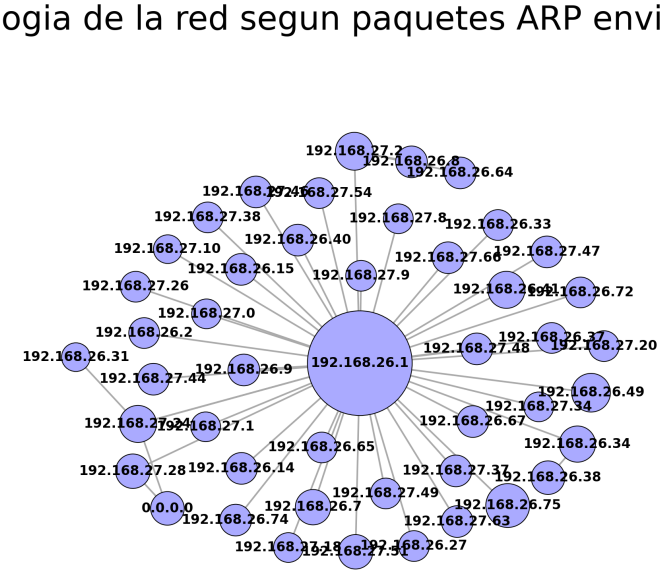

Figura 4

En el centro del gráfico y de mayor tamaño podemos observar el nodo que representa al router con dirección IP 192.168.26.1. Por destacarse frente a los demás nodos creemos que este podría ser el nodo distinguido.

Vemos ahora el gráfico que muestra la cantidad de información de cada nodo con respecto a la entropía.

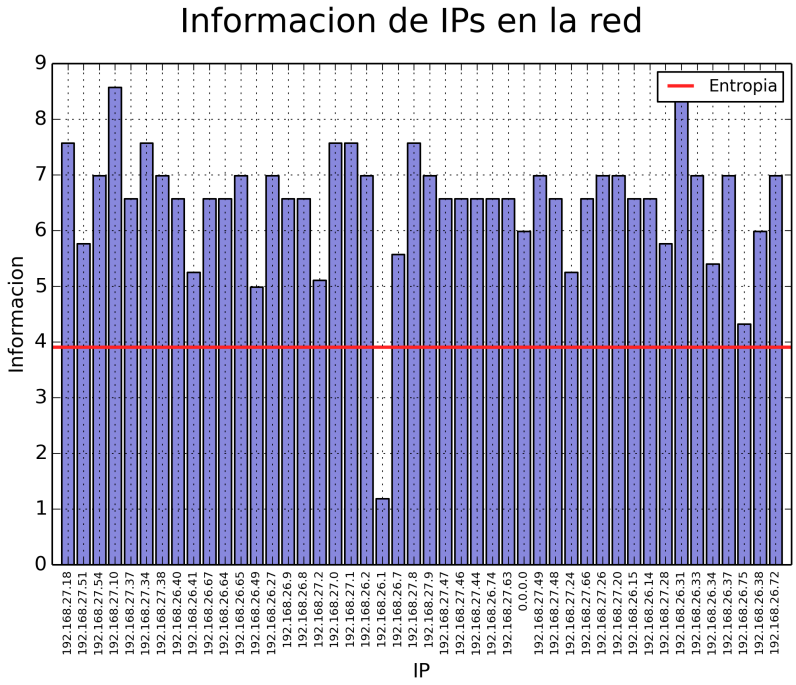

Figura 5

Como podemos comprobar en este gráfico, la información que presenta el nodo correspondiente al router se encuentra muy por debajo de la entropía de la fuente. Por lo tanto concluímos que este es un nodo distinguido. Por otro lado vemos a un par de nodos que aportan mucha información, o sea que presentan una baja probabilidad. Posiblemente estos dos nodos sean equipos que tuvieron poca actividad durante la captura de paquetes, y participaron en pocas búsquedas ARP.

Veremos a continuación el gráfico que muestra la cantidad de información de cada protocolo en comparación con la entropía.

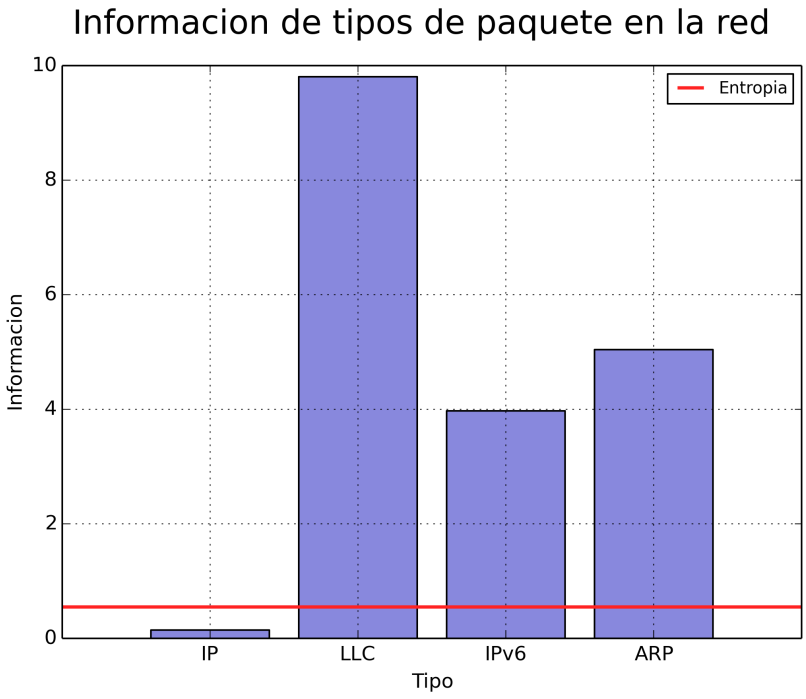

Figura 6

Se pueden observar dos símbolos distinguidos. El protocolo IPv4, cuya información se presenta por debajo de la entropía. Es el protocolo que presenta más frecuencia y por este motivo aporta muy poca información. Y el protocolo

LLC que con muy poca frecuencia nos da mucha información. También podemos observar que la cantidad de paquetes ARP e IPv6, en comparación con la de IPv4, es baja.

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 detalles de la naturaleza de los equipos que intercambian información en la misma. Pero conjeturamos que la mayoría podrían ser teléfonos celulares.

Vemos a continuación el gráfico que muestra la topología según la captura.

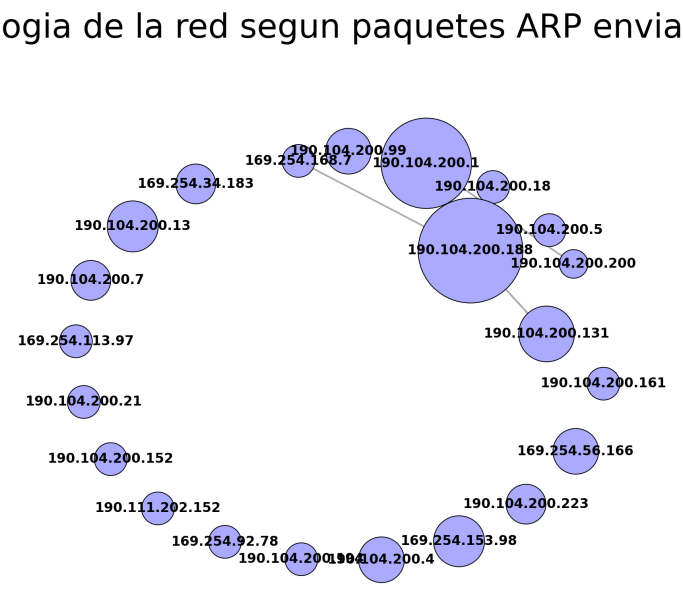

Figura 7

En este diagrama podemos observar que la mayoría de los nodos se encuentran sin un eje que los conecte. Esto se debe a que durante la captura, en general los paquetes ARP tuvieron la misma dirección IP como origen y destino. Según lo que investigamos sobre este comportamiento se debe a una práctica llamada ARP gratuitos. Los paquetes ARP gratuitos contienen en el destino la misma dirección IP que la del origen quien lo envía y la MAC destino es una dirección broadcast. La utilidad de este tipo de paquetes es la de poder detectar conflictos de direcciones IP duplicadas. También se utilizan para informar a los switches, o máquinas, de las MAC de sus interfaces para que actualicen sus tablas sin necesidad de peticiones directas.

Observamos a continuación el gráfico de cantidad de información de cada nodo con respecto a la entropía.

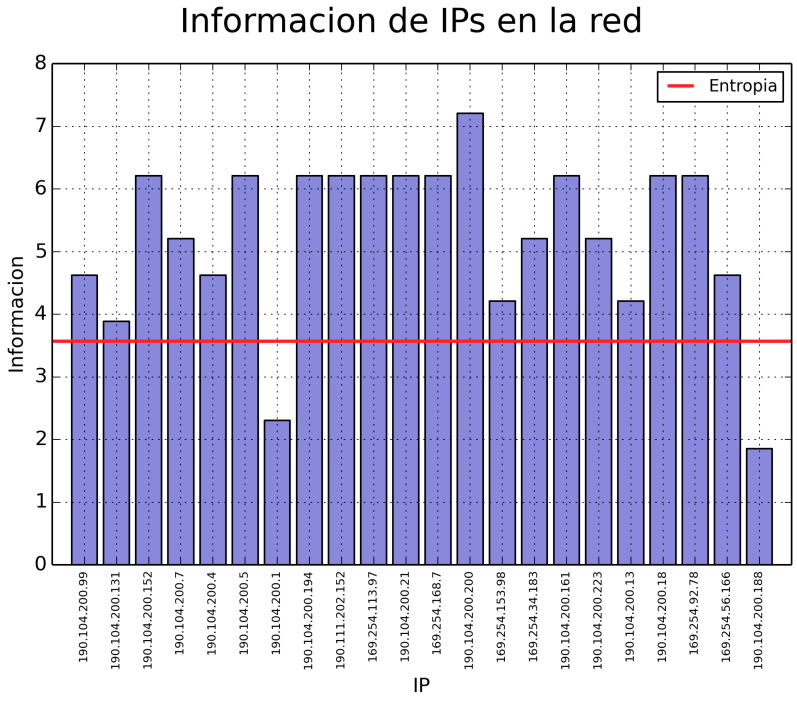

Figura 8

Como podemos observar, hay dos nodos que presentan una cantidad de información que se encuentra por debajo de la entropía. Basándonos en los experimentos analizados anteriormente, inferimos que alguno de estos dos nodos posiblemente sea el router.

Vemos el gráfico de la cantidad de información de cada protocolo en comparación con la entropía.

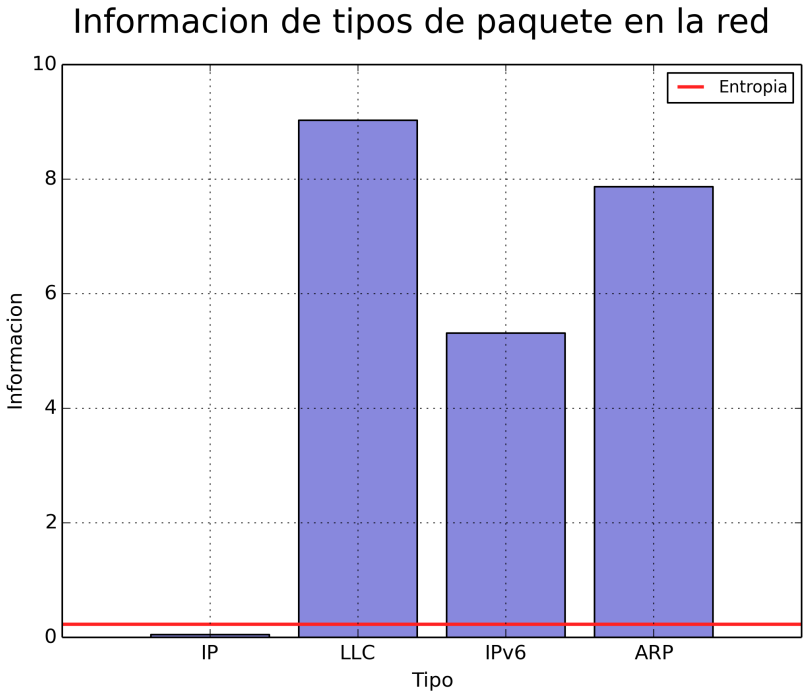

Figura 9

En este caso los resultados son similares a los obtenidos en los experimentos anteriores. Donde el protocolo IP presenta una cantidad de información por debajo del valor de la entropía. El resto de los protocolos, LLC, IPv6 y ARP, aparecieron con menor frecuencia, aportando mayor cantidad de información.

5. Conclusiones

En estos experimentos pudimos apreciar una aplicación concreta de la teoría de la información que nos permitió modelar y analizar diferentes fuentes de información.

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 el análisis de protocolos pudimos encontrar muchas similitudes entre las distintas redes. En general la entropía de esta fuente presentó un valor bajo y además el protocolo IPv4, que es el que mayor tráfico presenta, aportó valores de información inferiores. Esto demuestra una fuente predecible en la que se espera que la mayoría de los paquetes pertenezcan al protocolo IPv4.

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 pudo observar que en las redes analizadas el router fue uno de los nodos distinguidos.

También se observó que para las fuentes que modelan los diferentes nodos de la red, el valor de la entropía fue más elevado que la entropía de las fuentes que modelan los tipos de protocolos. Esto nos parece razonable debido a la impredecibilidad de la incidencia de los nodos en la red en comparación con la de los diferentes tipos de protocolos.