DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 1

Wiretapping

7 de diciembre de 2015 — Teoría de las comunicaciones

Integrante	LU	Correo electrónico
González, Sergio Martín	723/10	sergiogonza90@gmail.com
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. Introducción

El objetivo de este trabajo, es poder realizar un estudio sobre distintas redes, para poder realizar un análisis de los distintos **protocolos** que corren en la misma, así como la posible caracterización de algunas IPs, como **nodos distinguidos**

Para realizar esto, se implemento una herramienta capas de: poder capturar el trafico de una red, ya sea Wifi o Ethernet 802.3, y quedarse solo con la información de los distintos protocolos encontrados, y los distintos hosts (IP/MAC) que intervienen.

La herramienta también permite calcular la entropía, y la probabilidad de aparición de un nodo, en el trafico de la red en un período de tiempo determinado.

Con esta información, el objetivo es realizar un análisis exhaustivo de las redes estudiadas.

2. Primera Consigna: Caputando tráfico

2.1. Fuentes de información

A continuación se presentan las fuentes de información del modelo teórico, que luego se simularan mediante la herramienta implementada.

2.1.1. Protocolos

En primer lugar, la cátedra planteo una fuente de información que nos ayude a distinguir entre los distintos protocolos que se pueden hallar en una red. La fuente de información es la siguiente:

$$S_{t_i;t_f} = \{s_1 \cdots s_n\} \text{ siendo } s_i = p_i.type / p_i \in P \text{ entre los instantes de tiempo } [t_i; t_f].$$

y P esta definida de la siguiente manera:

$$P_{t_i;t_f} = \{p_1 \cdots p_n\} \text{ siendo } p_i \text{ el i-esimo paquete transmitido en la red entre los instantes de tiempo } [t_i, t_f]$$

Entonces, lo que se hace es, realizar una captura de la red, entre los instantes t_i y t_f, y de esto nos quedamos únicamente con los campos *type* del frame Ethernet. Así obtenemos una fuente de información donde cada símbolo es un protocolo utilizado en la red.

2.1.2. Nodos

Ahora, debemos idear una fuente de información, que nos permita poder distinguir los hosts involucrados en la captura de trafico, y a su ves, distinguir entre ellos, cuales son los mas concurridos y con mas apariciones en la red.

Para esto, tomemos la fuente de información P vista en el anterior punto, y tratemos de adaptarla para poder obtener una fuente de información que nos permita encontrar hosts en la red, solo basándonos en paquetes ARP. Tomemos el siguiente subconjunto de P:

$$\overline{P}_{t_i;t_f} = \{p_1 \cdots p_n\} \; \forall \; p_i \in \text{P} \; / \; p_i.type = ARP \text{ entre los instantes de tiempo } [t_i; t_f].$$

Con esta fuente de información, lo que hacemos es quedarnos únicamente con los paquetes ARP de toda la captura de trafico. En base a esto proponemos la siguiente fuente de información.

$S_1 = R_{t_i;t_f} = \{r_{ai} \mid r_{ai} = \overline{p}_i[ARP].ip\_origen\} \bigcup \{r_{bi} \mid r_{bi} = \overline{p}_i[ARP].ip\_destino\}$ entre los instantes de tiempo $[t_i; t_f]$.

Esto quiere decir, nos quedamos con las IPs origen y destino de la capa ARP del paquete (del cual sabemos que tiene una, porque los símbolos son tomados de $\overline{P}$). De esta forma, obtenemos una fuente de información donde cada símbolo es un hosts de la red.

Teniendo esto, podemos calcular, tanto la entropía de ambas fuentes, así como la probabilidad y cantidad de información de cada símbolo, tanto para S_1 como para $S_{t_i;t_f}$. La cantidad de información de cada símbolo s, se calcula de la siguiente forma:

$$I(s) = -log_2(P(s))$$

en donde $P(s)$ es la probabilidad de ocurrencia de s, en todo el espacio muestral S, en este caso, todos los símbolos emitidos por S_1 y $S_{t_i;t_f}$. Luego, la entropía de la fuente es la que se obtiene mediante la siguiente formula:

$$H(S) = \sum_{s \in S} P(s) * I(s)$$

2.2. Resumen de la herramienta

Para la simulación de las fuentes de información anteriormente presentadas. Desarrollamos una aplicación utilizando Scapy, la cual escucha pasivamente la red, captura los paquetes y vuelca la información pertinente en una serie de archivos. Los mismos son los siguientes:

- Un archivo que indica: El origen y el destino (MAC) de la capa Ethernet, el "type", y las IPs origen y destino (solo cuando es pertinente)
- Un archivo con la probabilidad de ocurrencia de cada "type", y la entropía de los "type"
- Un archivo con la probabilidad de ocurrencia de cada "host" (ip), y la entropía de los host (solo teniendo en cuenta paquetes ARP)
- Una imagen con un grafo indicando los "request" (who-has) y "replys" (is-at) que se enviaron los "host" entre si (opcional).

2.3. Ejecución

La ejecución del a herramienta debe ser realizada en un entorno Linux, y es necesario tener instalado: *Python 3.0* (o superior), *Scapy*, y si ademas se quiere realizar grafos, es necesario tener instalado *graphviz*.

Para ejecutar la herramienta, se debe abrir una consola en la carpeta "src" adjunta a este informe y ejecutar el comando:

```
$ sudo ./WiretappingTool.py
```

Esto comenzara a capturar, y volcara la información en los archivos configurados como default. Adicionalmente, la aplicación cuenta con ciertos paramentos para personalizar la ejecución de la captura. Por ejemplo:

```
$ sudo ./WiretappingTool.py -f salida.out -t 60 --console --arp
```

Esto realiza lo siguiente:

- (−t o −−timeout) indica un tiempo en segundos para finalizar la captura, en esta caso 60 segundos.
- (−−arp) Captura solo paquetes ARP.
- (−−console) Ademas de volcar el resultado de la captura a un archivo, se muestran los paquetes en la consola, en tiempo real.
- (−f) Vuelca la captura en *salida.out* en lugar del archivo configurado por defecto (out/sniff.out).

Si ademas de la captura, se desea realizar un grafo con los nodos de la red, junto con los "request" y "replys" realizados por los mismos, se debe pasar el parámetro "–graph". Si se hace esto, al terminar la ejecución, la imagen con el grafo se abrirá al instante. Para ver con detalle cada uno de los parámetros que dispone la aplicación, se puede utilizar:

```
$ sudo ./WiretappingTool.py -h
```

3. Segunda Consigna: Gráficos y Análisis

Para poder simular las fuentes de información antes mencionadas, lo que hicimos fue correr la herramienta implementada en diferentes redes con distintas topologias. Todas las capturas fueron realizadas en redes Wifi, ya que en redes Ethernet "switcheadas" no se pueden ver los mensajes ARP *"is-at"* que no son hacia nuestro host, debido a que estos son unicast, y no es posible capturarlos debido a que no llegan a la interfaz de la pc que realiza la captura.

En primer lugar, queremos presentar las redes que utilizamos, y ademas mostrar el grafo que genero nuestra herramienta, con los mensajes ARP "*who-has*" y "*is-at*"

3.1. Red Laboral

Esta red es la red del trabajo de uno de los integrantes del grupo. Dicha red cuenta con diversos routers Wifi, por lo que es difícil saber de cuantos hosts es posible capturar el trafico. Por los datos obtenidos se puede estimar que al momento de realizar la captura, se encontraban conectados alrededor de 53 hosts. A continuación se muestra el grafo resultante de la captura.

Como se puede ver a simple vista en la figura 1, hay 3 IPs muy concurridas: 192.168.28.139 (A), 192.168.28.143 (B), y 192.168.29.254 (C). Debido a que la IP (C) termina con el numero 254, y que ademas es el nodo mas concurrido de la red (al momento de esta captura), podemos intuir que esta IP es el router al que esta conectada la PC que capturo el trafico. Luego, esta captura se realizó desde una maquina virtual corriendo Linux, hosteada en una pc con Windows 7, y las IPs (B) y (A) son el host real y la Virtual respectivamente.

3.2. Red Starbucks - FibertelZone

La siguiente captura se realizo también desde una maquina virtual corriendo Linux, desde una notebook corriendo Windows 8.1, en una red (Wifi) de Starbucks. De todas maneras, al momento de conectarse, la red Wifi pertenecía a la red de FibertelZone, sin contraseña, por lo que es muy probable que los hosts no se encuentren solo en el establecimiento, sino también en los alrededores.

Hay varias cosas a destacar en el grafo. Primero hay 2 redes visibles, una con IPs privadas (10.0.0.0) (A), y otra con IPs publicas de la red 169.254... (B) (no es posible determinar la mascara debido a que no hay suficientes IPs). Otra cosa a destacar, que es bastante interesante, es que hay paquetes ARP "who-has" desde IPs de la red (A) hacia una IP que podría legar a ser la IP broadcast de la red (B). Si esto fuera así, habría hosts preguntando por la dirección MAC de una IP broadcast, lo cual es bastante absurdo. Ademas el hecho de que esta IP no responda los "who-has", aumenta la posibilidad de que sea una IP broadcast. De todas maneras, al no saber la mascara de esta red, esa IP podría no ser broadcast. Por ejemplo si la mascara fuera 255.254.0.0, entonces la IP 169.254.255.255 seria una IP utilizable por algún host o un router, y no una IP broadcast.

Luego, también se puede visualizar paquetes "who-has" desde la ip 0.0.0.0, en ambas redes. Estos paquetes ARP, luego de investigar un poco, son denominadas Gratuitous ARP[1], las cuales sirven (entre otras cosas) para que un host verifique que no haya otro dispositivo usando su propia IP.

[1] https://wiki.wireshark.org/Gratuitous_ARP

Figura 1: Red Laboral

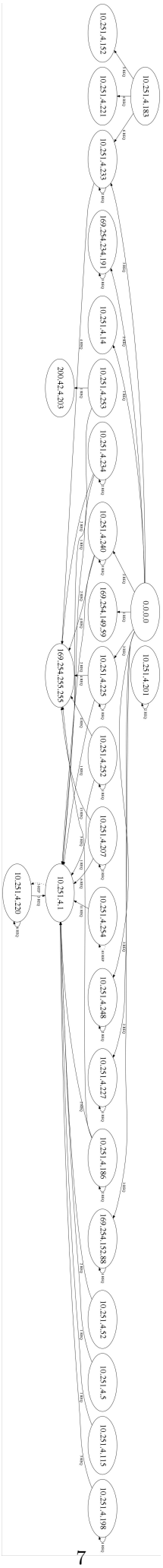

Figura 2: Red Starbucks - FibertelZone

3.3. Cantidad de información por nodo

El primer análisis que queremos desarrollar, es sobre la cantidad de información que provee cada nodo de la red. Para esto vamos a utilizar la fuente de información de nodos (S_1) y vamos a realizar un histograma mostrando la cantidad de información de cada IP.

Para poder identificar si un host nos provee "mucha" o "poca" información, necesitamos mostrar también la entropía de la fuente. Esto lo vamos a indicar con una linea horizontal con el valor de la misma. De esta manera, lo esperable sería que los hosts a los que se hacen muchos requests ARP, no sean nodos que en si provean demasiada información, ya que es algo habitual que estos se demanden.

3.3.1. Red Laboral

Esta es la red en la que se observó una mayor cantidad de hosts, y como se puede ver en la figura 3, hay 2 nodos distinguidos en la red, ya que estos son los que mas se observaron en el trafico, y por eso es que su nivel de información es bajo.

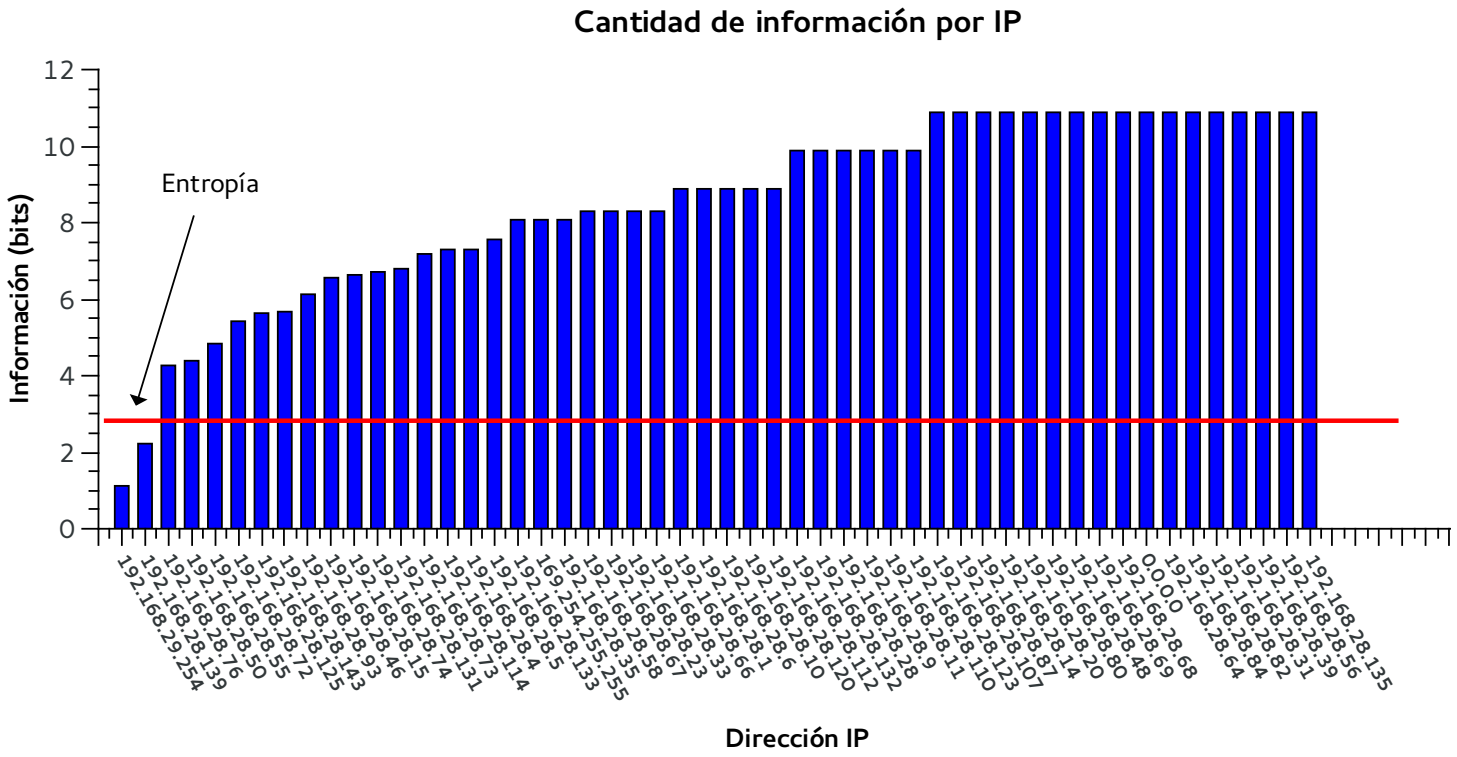

Figura 3

Según vimos en la sección anterior, la IP 192.168.29.254 es la IP del router al cual esta conectada la PC host de la virtual que realizo la captura, y la IP 192.168.28.139 es efectivamente esta virtual. Sabiendo esto, podemos pensar que como en realidad ambas IPs son de total conocimiento por el host que capturo el trafico, es razonable que la cantidad de información que nos proveen, estén por debajo de la entropía de la fuente.

3.3.2. Red Starbucks

En esta red se observo una menor cantidad de nodos que en la anterior. En esta, hay varias IPs que caen por debajo de la entropía de la fuente, pero sin duda el nodo mas distinguido es 10.251.4.1

También vemos como efectivamente la IP 169.254.255.255 aparece también como un nodo distinguido de la red, pero por lo que vimos antes, el echo de que no envíe paquetes "is-at", y de que termine con 255.255, hace muy probable que esta sea la dirección broadcast de otra Red. Luego también se ve una IP que no pertenece a las 2 redes que mas vemos en el histograma, la IP 200.42.4.203. Esta IP es la de mayor incertidumbre de la red, ya que es la IP con mayor información de la captura, y esto lo hace estar muy por encima de la entropía de la fuente.

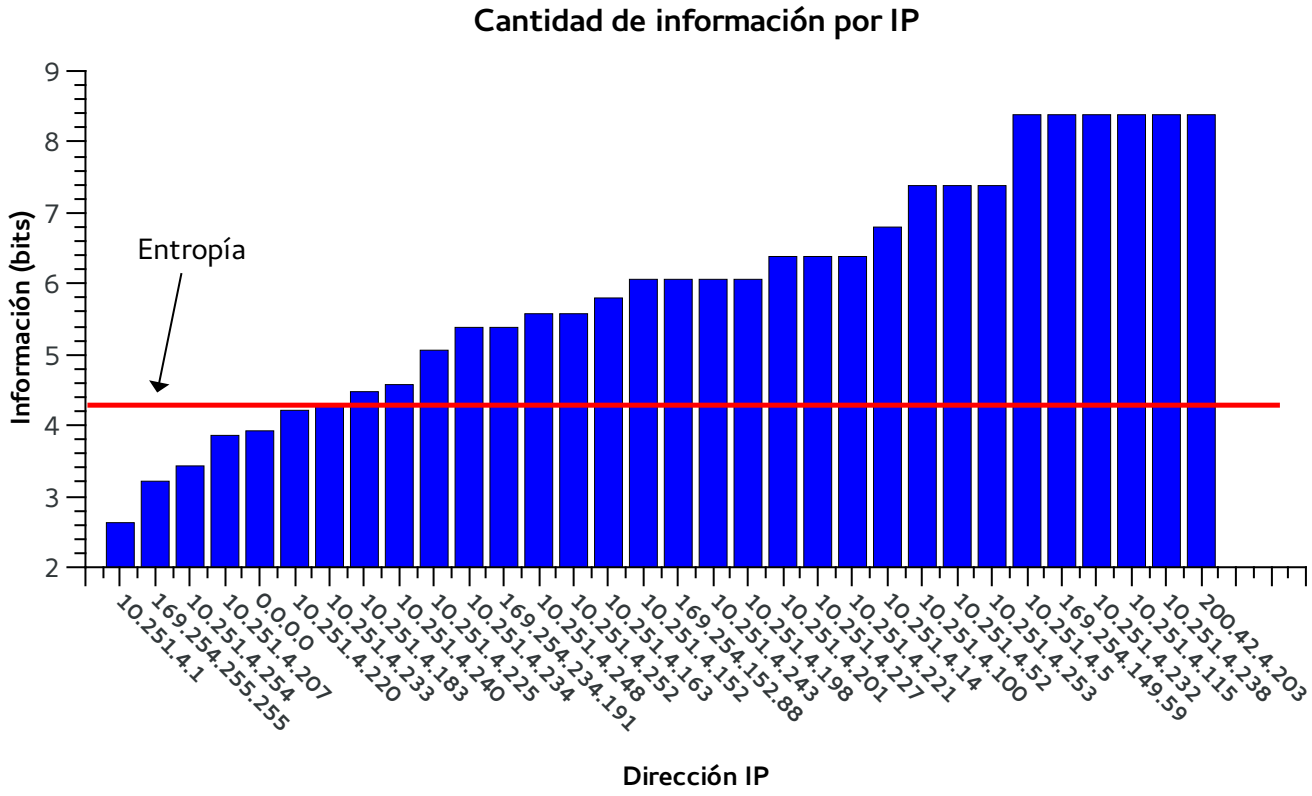

Figura 4

3.4. Paquetes capturados de cada protocolo

El siguiente paso es observar un poco la simulación de la fuente de información que fue provista por la cátedra, y esto es, ver los distintos protocolos que son visibles a través del campo *type* de los paquetes Ethernet.

En otras capturas que realizamos se observaron algunos paquetes LLC, pero en las capturas que luego decidimos estudiar, no se observaron paquetes de esta índole, y solo se vieron 3 protocolos: ARP, IPv4 y IPv6. La idea entonces, es mostrar en un gráfico de torta, cual fue la probabilidad de cada símbolo (protocolo) de esta fuente de información.

3.4.1. Red Laboral

En la figura 5, se ve como los paquetes ARP casi son un 10 % del total, e incluso es mayor al porcentaje de paquetes IPv6 de la red. Quizá esto sea debido a que el tiempo de expiracion de las entradas de las tablas ARP esten configurados con un tiempo muy chico, o que haya demasiados cambios en la red y esto provoque que los hosts deban continuamente estar averiguando la IP del router o de algún otro host.

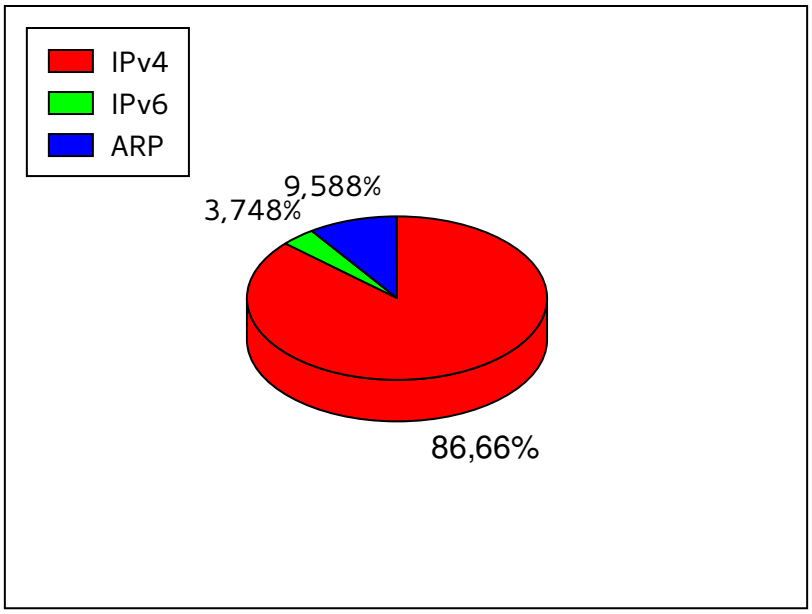

Figura 5

3.4.2. Red Starbucks

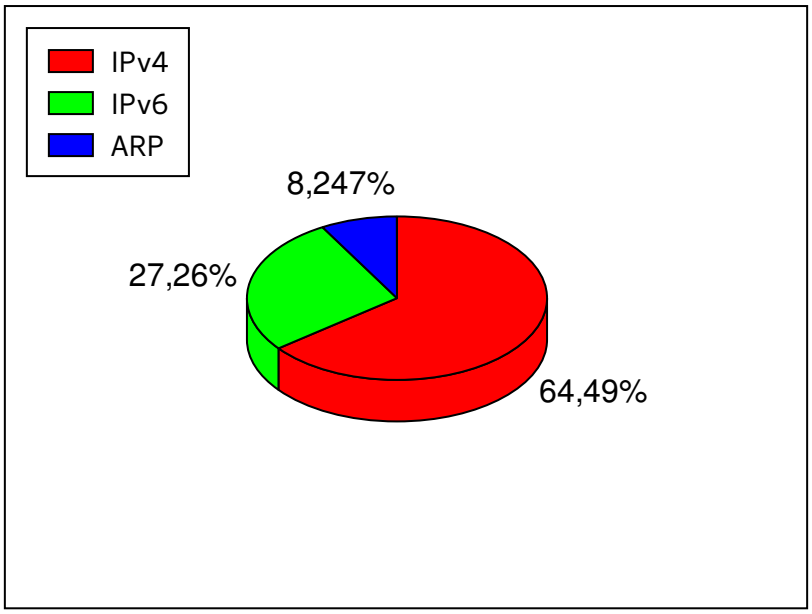

Figura 6

En este caso el porcentaje de paquetes ARP es un poco menor, y aumenta mucho mas el porcentaje de paquetes IPv6.

3.5. Variación de la entropía a lo largo del tiempo

El siguiente análisis es bastante interesante, debido a que nos da una idea de como varía la entropía a lo largo del tiempo, y de que tan grande debe ser una captura para poder saber un valor certero de la entropía de la red. También hay que tener en cuenta que la entropía puede variar a lo largo del dia, ya que el trafico en internet es muy fluctuante, y el movimiento posiblemente no sea el mismo en horas de la madrugada que a las 3 de la tarde.

Par realizar este análisis optamos por mostrar un gráfico de lineas en donde se muestra la variación de la entropía en función del tiempo, para ambas fuentes de información. Cabe aclarar que el hecho de mostrarlas en el mismo gráfico

es por una cuestión de claridad, y de poder compara el nivel de insertidumbre de una fuente con otra, y ademas el poder comprarlas con respecto a cuanto tardan en converger a un valor estable.

3.5.1. Red Laboral

Para la red laboral, se puede ver como la entropía para la fuente de información de nodos, se estabiliza luego de pasados los 9 minutos de captura. En cambio la fuente de protocolos, lo hace a los 7 minutos, y alrededor de los 3 minutos presenta una variación muy chica. Esto quiere decir que, los protocolos en si suelen presentar un nivel de incertidumbre muy constante porque ya son algo intrínseco de la red, y no son muy fluctuantes. En cambio, en cuanto a los nodos, al ser una red wifi donde suele haber mucho movimiento de notebooks y de otros dispositivos, la fluctuación es mayor, y esto hace la entropía tarde mas en converger.

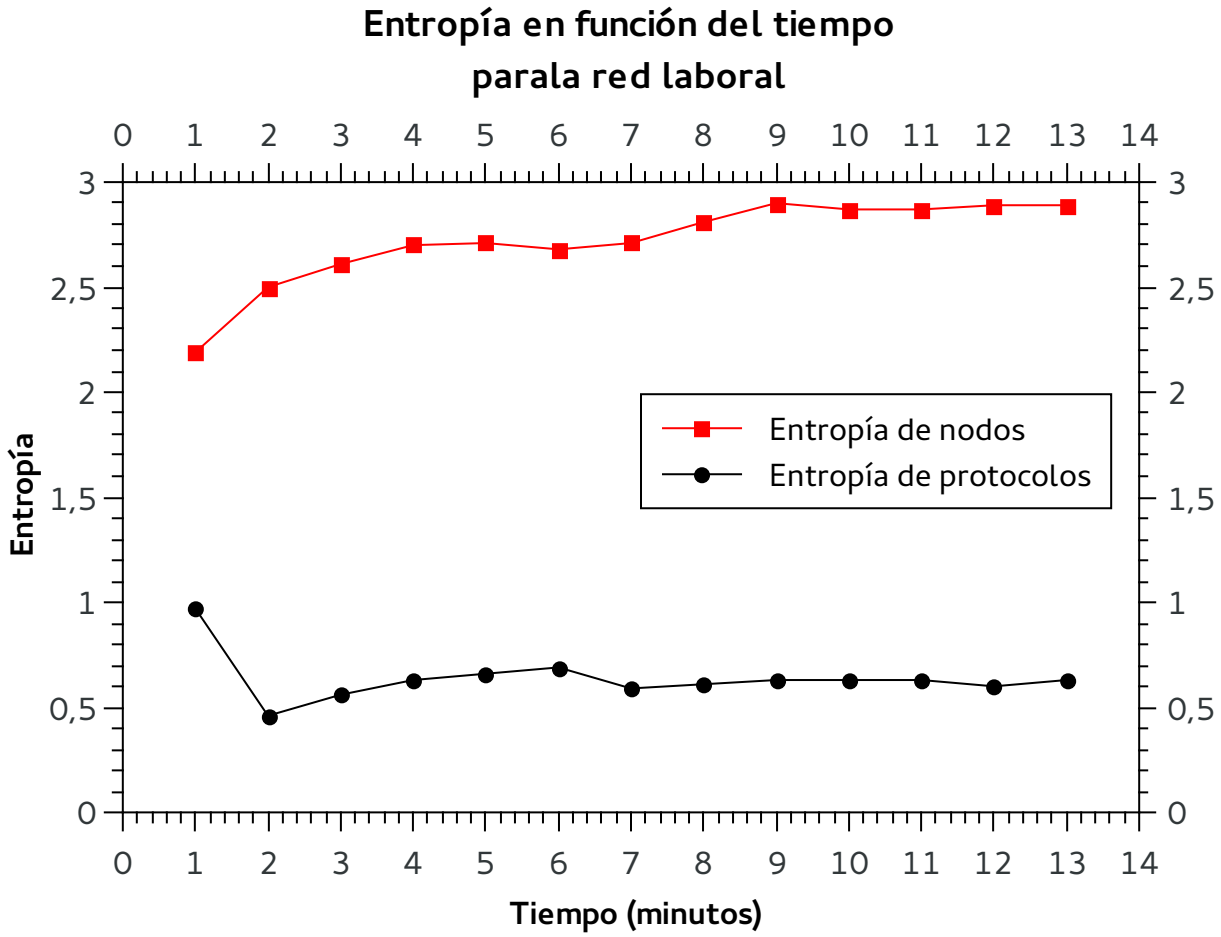

Figura 7

3.5.2. Red Starbucks

En esta red, pasa algo parecido que en la red anterior, ya que la fuente de información de nodos tarda mucho mas en converger que la de protocolos, pero acá es mucho mas notable como la fluctuación de los nodos hace que varíe mucho la entropía de los nodos a medida que pasa el tiempo, ya que esta comienza siendo de 2.75 bits, y termina siendo de 4.25. En cambio en la entropía de la fuente de información de protocolos, esto es mucho menor, y casi es constante desde el comienzo de la captura.

También se puede ver como en general, en ambas redes la entropía, osea, el nivel de incertidumbre de la fuente, es bastante mayor en cuanto a nodos que en cuanto a protocolos. Esto también es por lo dicho anteriormente, de que los protocolos ya son algo intrínseco de las redes, y no hay tanta incertidumbre con respecto a eso, pero si la hay en cuanto a los nodos que se conectan a la misma.

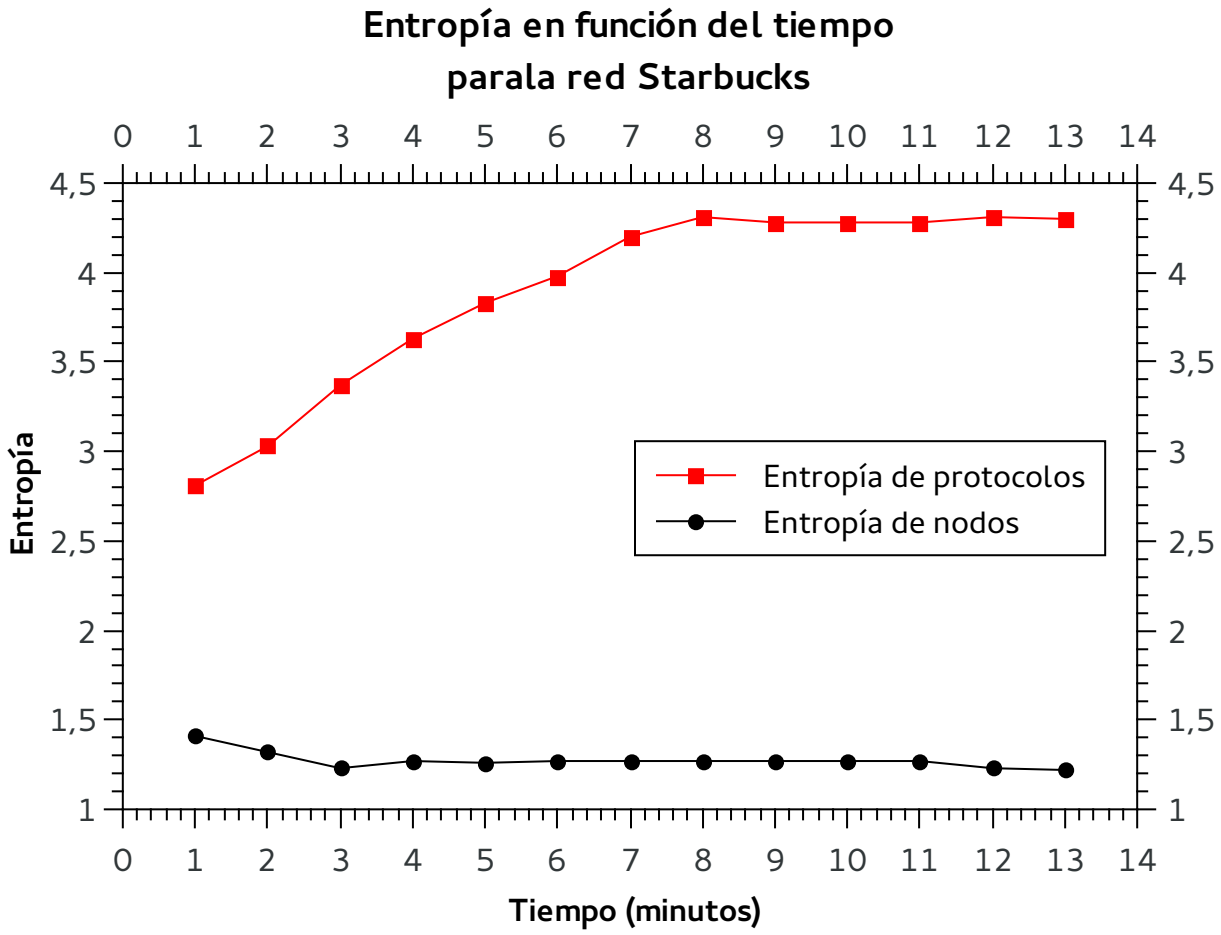

Figura 8

4. Conclusiones

En este trabajo, pudimos estudiar el comportamiento de 2 redes con una densidad de nodos bastante grande, y pudimos ver cuales son los nodos mas distinguidos y los protocolos predominantes de las mismas.

Vimos que mediante un simple análisis de la cantidad de información de cada nodo, se pueden obtener fácilmente los nodos mas concurridos. También vimos como el porcentaje de protocolos ARP no es tan pequeño en algunos casos, como por ejemplo en la red Laboral, en donde el porcentaje de paquetes ARP superaban al porcentaje de paquetes IPv6.

También pudimos hacer un análisis de cuanto tiempo es necesario capturar trafico de una red, para poder obtener una medida estable de la entropía. Se observo que la fluctuación de nodos en una rede Wifi puede ser muy grande, y esto provoca que la entropía varíe mucho, a diferencia de la entropía de protocolos, la cual suele ser bastante constante debido a que son algo intrínseco de las redes hoy en día. Debido a este motivo, también pudimos ver como la entropia en si de los distintos nodos conectados a una red, es mucho mayor a la entropía de los protocolos que corren en la misma, ya que estos presentan mucha menor incertidumbre.