DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Recuperatorio de Trabajo Práctico I

Segundo cuatrimestre 2016 — Teoría de las comunicaciones

Integrante	LU	Correo electrónico
Oller, Luca	667/13	ollerrr@live.com
Zamboni, Gianfranco	219/13	gianfranco376@gmail.com
González, Rodrigo	294/01	rodrigogk@gmail.com

Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria - (Pabellón I/Planta Baja)
Intendente Güiraldes 2160 - 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 práctico fue aplicar técnicas provista por la Teoría de la Información para distinguir distintos aspectos de la red de manera analítica. Para ello, nos basamos en la captura pasiva de paquetes ARP dentro de una red para poder analizar posteriormente los datos obtenidos. Fue necesario entonces desarrollar una herramienta que permita obtener los paquetes ARP de una red. Luego fue necesario realizar una serie de experimentos sobre redes reales. Una vez obtenidos esos datos, se procedió a su procesamiento y análisis mediante cálculos y confección de gráficos.

2. Experimentación

2.1. Implementación de herramientas

Para poder llevar a cabo de manera satisfactoria los experimentos, se implementó una herramienta que escuche pasivamente los paquetes de una red, calcule las características destacadas de la misma (entropía, información los símbolos, etc) y luego los organice y muestre de manera adecuada. De está forma, se facilitó, además, la comparación y procesamiento de dichos datos y se pudo llegar a conclusiones de una manera rápida e intuitiva.

Dicha herramienta, dependiendo los parámetros pasados realiza las siguientes acciones:

1. Escucha los paquetes ARP de la red a la que se encuentra conectado el dispositivo durante 15 minutos o levanta un archivo de paquetes de red con formato libpcap
2. Por cada paquete conseguido, calcula la entropía de la red y la cantidad de información que proporciona cada símbolo de la fuente modelada.
3. Cuando todos los paquetes fueron analizados, imprime por pantalla una matriz de valores, siendo la primer columna la entropía y el resto los valores de información de cada símbolo. Con este resultado, se intenta mostrar como variaron las distintas propiedades de la red y de los símbolos con cada paquete enviado. Así, la i-ésima fila de la matriz representa el estado de la red cuando llegó el i-ésimo paquete.
4. Adicionalmente se cuenta con una herramienta que permite realizar un gráfico de la red subyacente e histogramas de cantidad de paquetes y cantidad de información, para el caso de que los símbolos de la fuente se asocien con las diferentes IP de la red.

2.2. Las fuentes modeladas

Este trabajo se basa en el análisis de los datos obtenidos en experimentos sobre redes de diferentes características. Luego, los paquetes capturados son modelados como una fuente de memoria nula. En este caso, se adoptan dos modelos diferentes:

La fuente S como una fuente binaria cuyos símbolos clasifican cada paquete en *Broadcast* o *Unicast* dependiendo si el destino de los mismos es la direción MAC $ff:ff:ff:ff:ff:ff$ o no lo es, respectivamente.

La fuente $S1$ que considera únicamente los paquetes ARP de tipo who-has. Realizando la distinción de los símbolos por las direcciones IP de origen de estos paquetes. La elección del modelo de la fuente $S1$ se realizó luego de probar diferentes alternativas:

- Origen de todos los paquetes ARP
- Destino de todos los paquetes ARP
- Origen de los paquetes ARP del tipo is-at
- Destino de los paquetes ARP del tipo is-at
- Origen de los paquetes ARP del tipo who-has
- Destino de los paquetes ARP del tipo who-has

En base a los resultados obtenidos, encontramos que el modelo de fuente seleccionado nos permitía distinguir la mayoría de los nodos de la red subyacente (no aparecía el mismo conjunto de nodos si se seleccionaba la dirección de origen o destino de cada tipo de paquete). Además, permitió obtener valores consistentes en cuanto a cantidad de paquetes que uno esperaba para la IP correspondiente al default gateway en la mayoría de los experimentos realizados.

Cabe aclarar que cuando mencionamos *destino* nos referimos al *target* del paquete ARP (es decir, la dirección por la cual se consulta en los mensajes who-has o la dirección a la que se responde en los mensajes is-at).

2.3. Experimentos

Para la experimentación se realizó la captura de paquetes en tres redes distintas utilizando Wireshark. La duración de cada uno de los experimentos fue de 15 minutos.

- El primer experimento se hizo sobre la red LAN Ethernet de la empresa Aserpel SA, en horario laboral, de pleno uso de la red.
- El segundo experimento se realizó en el laboratorio del DC durante horas de la tarde.
- El tercer experimento se llevó a cabo sobre la red wifi de un local de comidas rápidas Wendy's en Belgrano. El local se encontraba bastante concurrido.

2.4. Gráficos

Para la fuente binaria S, se realizaron gráficos mostrando la variación de la entropía de la fuente y la cantidad de información de cada uno de los símbolos: broadcast y unicast. También es posible visualizar en el mismo gráfico el valor de la entropía máxima alcanzada a través del tiempo (en realidad, es en función de la cantidad de paquetes recibidos hasta el momento, pero puede considerarse equivalentes) y el valor de la entropía máxima teórica (es decir, alcanzada con la equiprobabilidad de los símbolos).

Por otro lado, presentaremos un gráfico mostrando la red de mensajes ARP subyacente.

Y por último, tendremos un gráfico de barras para la fuente $S1$ con la cantidad de información calculada para cada uno de los símbolos (es decir, las diferentes direcciones ip).

3. Resultados

3.1. Red Lan Aserpel

3.1.1. Cantidad de información dada la fuente binaria S

A continuación tenemos el gráfico que permite visualizar la variación de la entropía y la cantidad de información a través del tiempo. De acuerdo a las mediciones realizadas, la fuente S alcanzó un valor máximo de entropía de 0.22 bits. Un poco alejado del máximo teórico de toda fuente binaria (1 bit). Esto es así debido a que la fuente no es equiprobable. Notamos fácilmente ésto, observando la diferencia entre la cantidad de información provista por los mensajes broadcast contra la de los mensajes broadcast. Es decir, en el caso de esta fuente, son más recurrentes los mensajes unicast que los mensajes broadcast. Podemos decir que este comportamiento es esperable (mayor presencia de mensajes unicast), ya que los mensajes broadcast suelen provenir de protocolos de control y es muy raro ver datos de aplicación en mensajes que no sean unicast. Es deseable que el overhead impuesto por los protocolos de control sea bajo, y en este caso es así.

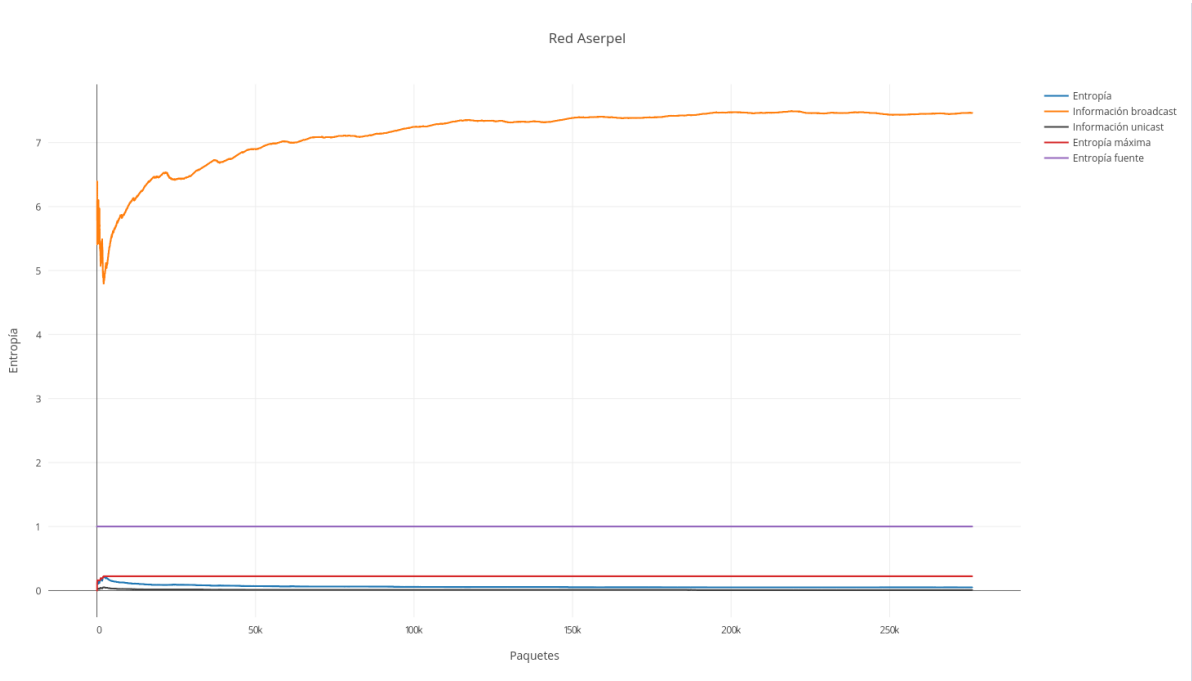

Figura 1: Entropía y cantidad de información de la red Aserpel según la fuente S

Para obtener una visualización más detallada de los valores de la entropía y de la cantidad de información de los mensajes broadcast, tenemos el siguiente gráfico ampliado.

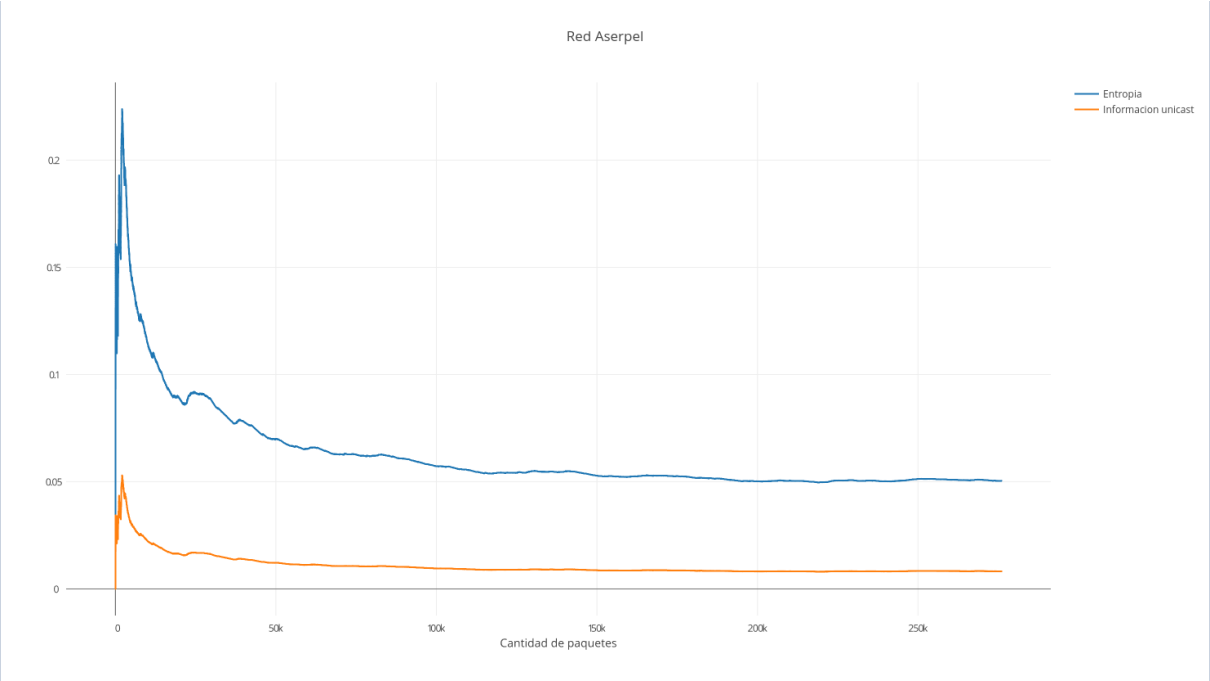

Figura 2: Entropía y cantidad de información de la red Aserpel según la fuente S (zoom)

Vemos que la cantidad de información provista por los mensajes unicast, es baja. De acuerdo a la teoría de la información, esto indica que es el tipo de mensajes que abundó durante el experimento.

Con respecto a la evolución en el tiempo, podemos notar que a partir de cierto momento los valores se estabilizan y tienden a volverse constantes.

3.1.2. Red de mensajes ARP subyacente

La red de mensajes ARP subyacente se puede ver a continuación:

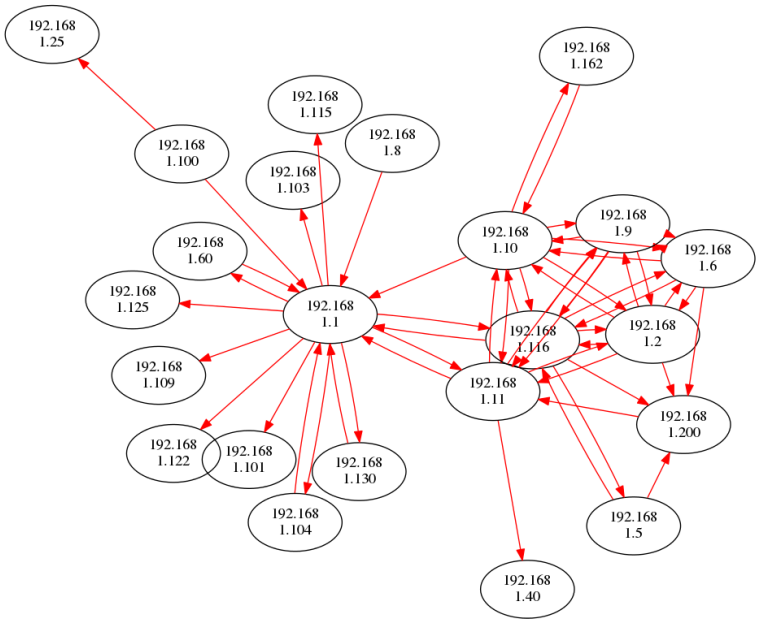

Figura 3: Aserpel - Red de mensajes ARP subyacente

Ésta es la red más pequeña con la que experimentamos en este trabajo (23 nodos según el grafo). Podemos notar la presencia de un nodo distinguido a simple vista: 192.168.1.1. Por el aspecto de la dirección IP, es posible que se trate del router de la red, ya que suele ser la dirección default utilizada.

3.1.3. Cantidad de información de la fuente S_1

Aquí tenemos el gráfico de cantidad de información aportada por cada dirección IP que aparece como origen de los paquetes $who - has$.

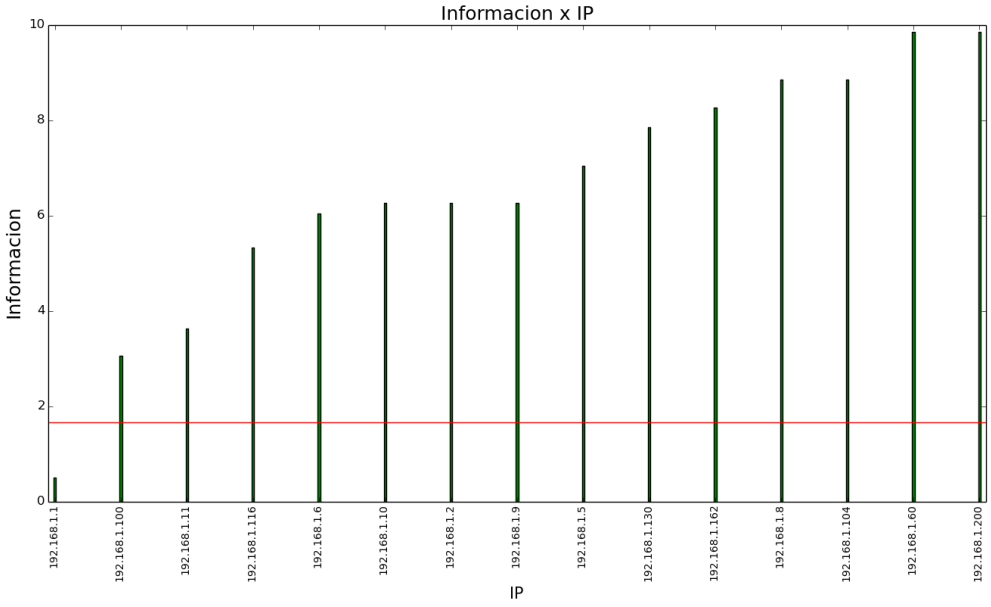

Figura 4: Cantidad de información de la red Aserpel de acuerdo a la fuente S_1

De acuerdo a los cálculos realizados, la fuente S_1 tiene una entropía de 1.82 bits. Podemos notar que hay un único nodo (192.168.1.1) cuya cantidad de información se encuentra por debajo del valor de la entropía de la fuente. Éste es el mismo nodo que sobresalía en el grafo de la red subyacente. De acuerdo a la teoría de la información, podemos decir que el nodo 192.168.1.1 es un nodo distinguido, ya que al aportar una cantidad baja de información y por debajo de la entropía, es emitido frecuentemente por la fuente S_1 (es decir, aparece muchas veces en los mensajes ARP considerados para esta fuente).

En el caso de la red de Aserpel, pudimos contrastar contra la realidad y la dirección IP 192.168.1.1 se trata efectivamente del default gateway. De esta manera, para este experimento, podemos decir que el método de considerar como nodo distinguido todo aquel cuya cantidad de información se encuentra por debajo de la entropía fue efectivo. Además, al tratarse de un único nodo pudo considerarse como default gateway indudablemente.

3.2. Red Lan Laboratorio DC

3.2.1. Cantidad de información dada la fuente binaria S

En la siguiente figura se muestra la evolución de la entropía de la fuente S y la cantidad de información de los mensajes broadcast y unicast.

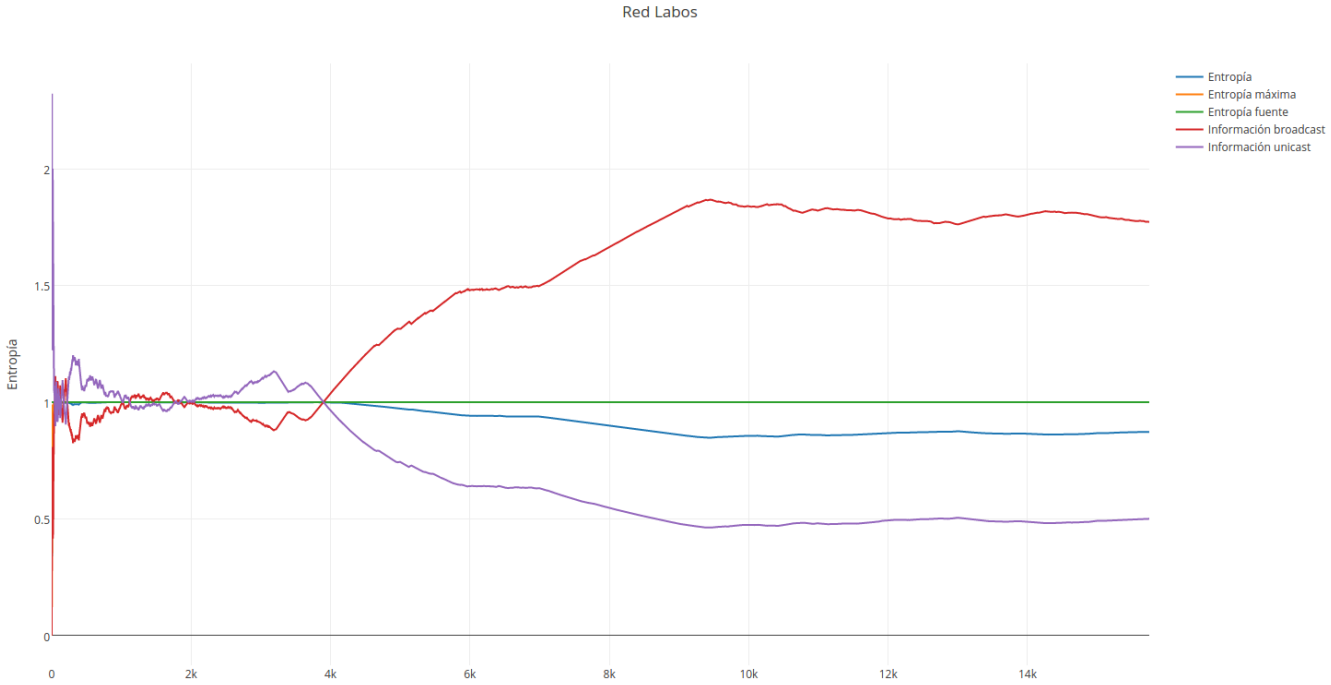

Figura 5: Entropía y cantidad de información de la fuente binaria S

Vemos que la cantidad de información de la fuente S hasta cierto momento se mantienen en valores cercanos, y den la idea de equiprobabilidad, alcanzando un valor máximo de entropía de 1 (esto ocurre porque en un momento "se cruzan"los valores de la cantidad de información del broadcast y unicast, en ese instante la entropía vale 1 y se alcanza el máximo). Podemos decir, que hasta el momento en que se reciben 4k paquetes, la red no tiene un comportamiento esperado, ya que la cantidad de información de broadcast es menor a la de unicast. Es decir, hay un alto overhead de los protocolos de control hasta ese instante. Luego, la situación se normaliza y se empiezan a obtener los valores esperados: cantidad de información del broadcast mayor a la cantidad de información de los mensajes unicast.

3.2.2. Red de mensajes ARP subyacente

A continuación, podemos ver la red de mensajes ARP subyacentes.

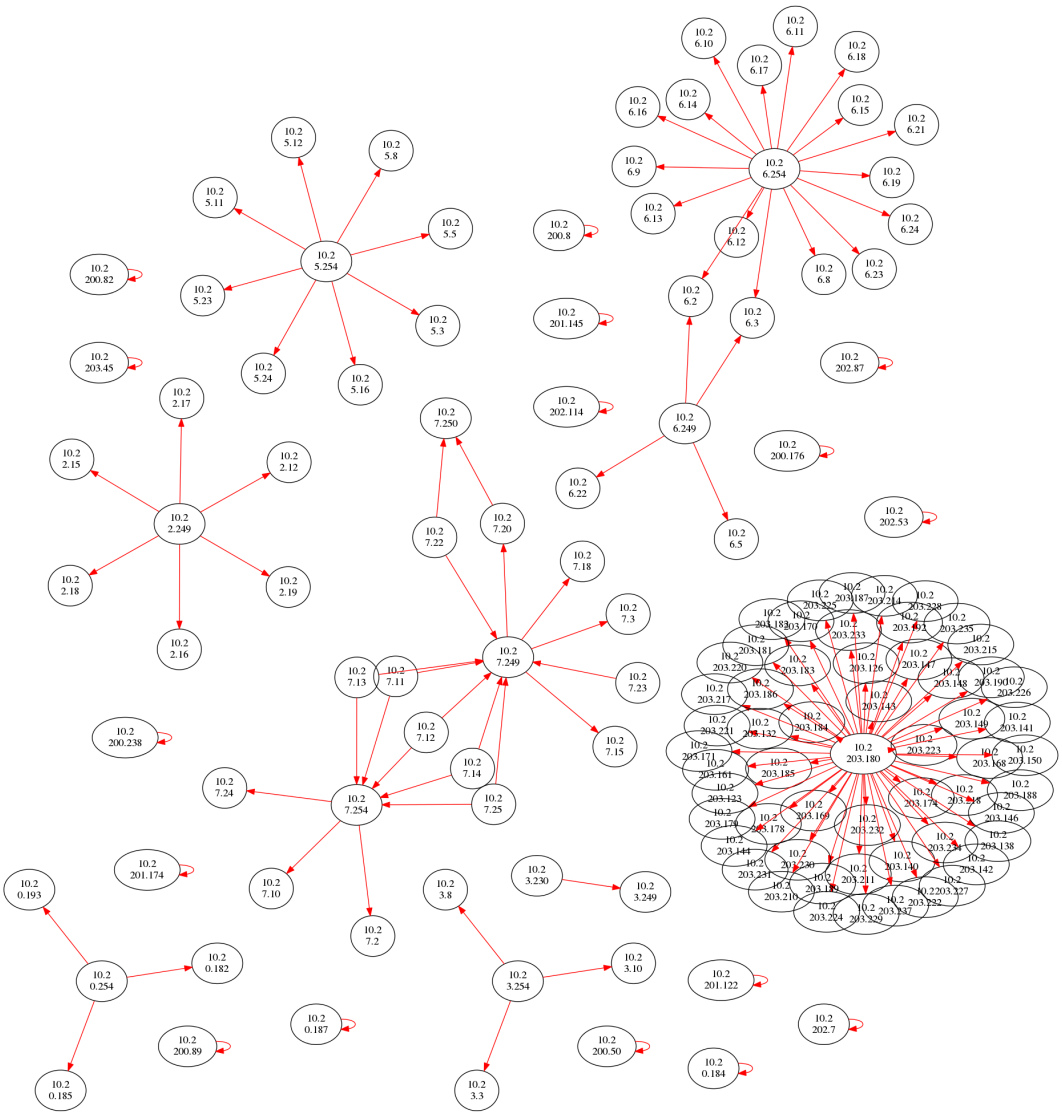

Figura 6: Laboratorio del DC - Red de mensajes ARP subyacente - Global

Esta puede considerarse como una red grande, ya que tenemos más de cien nodos participantes. Nos ha resultado bastante complejo su análisis. Trataremos de atacar algunas de sus particularidades.

Notamos un gran conjunto de nodos conectados a un nodo central (102.2.203.180). Luego, otros conjuntos con menor cantidad de nodos, interconectados por uno o varios nodos centrales. Además, nodos que se enviaron mensajes a sí mismos.

Podemos sospechar que el nodo 10.2.203.180 es el default gateway de la red, ya que es el que tiene mayor cantidad de nodos conectados. Por otro lado, los conjuntos más pequeños de nodos conectados, pueden tratarse de grupos pequeños de terminales que se conectan a otro, pero que no se conectan con el conjunto general (que podría suponerse que lleva a internet). Por último, los nodos que se apuntan a sí mismos, pueden estar apareciendo de esta manera, ya que uno de las aplicaciones posibles del protocolo ARP es el de la detección de direcciones IP duplicadas en la red y es común preguntar por la dirección IP de uno mismo (si otro nodo diferente a uno mismo contestara, significaría que la dirección IP se encuentra repetida en la red).

3.2.3. Cantidad de información de la fuente S_1

El gráfico de la cantidad de información de la red es el siguiente:

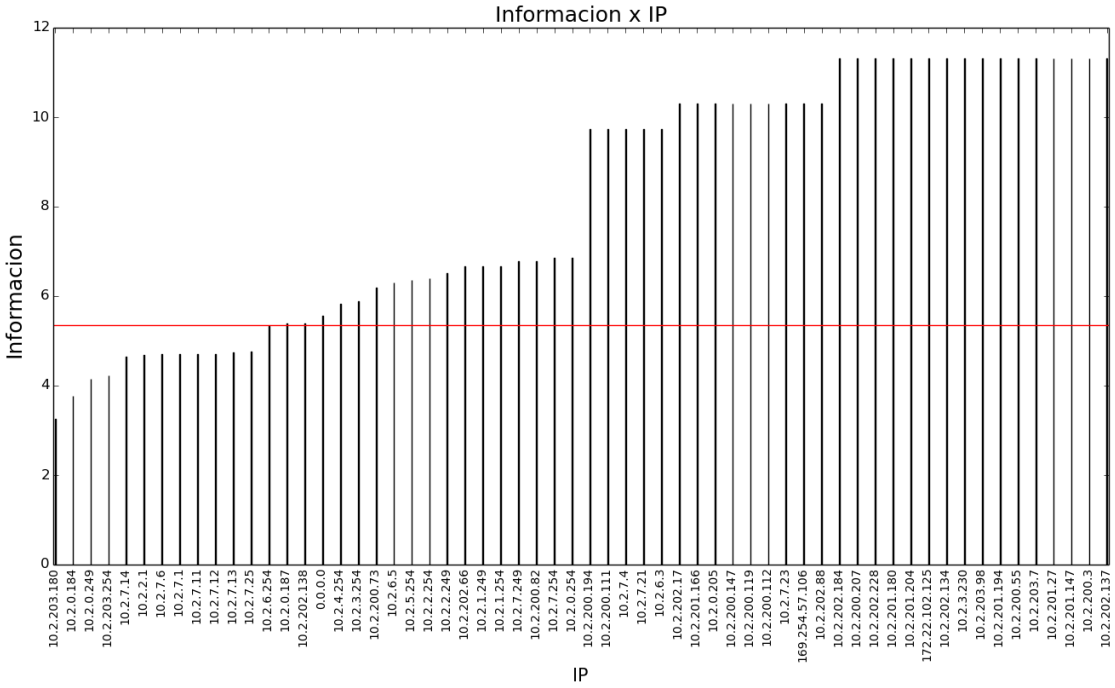

Figura 7: Cantidad de información de la red del laboratorio del DC respecto a la fuente S_1

La entropía de la fuente S_1 es de 5.22 bits. Los nodos cuya cantidad de información se encuentran por debajo de ese valor de entropía son varios (12 en total).

- 10.2.203.180
- 10.2.0.184
- 10.2.0.249
- 10.2.203.254
- 10.2.7.14
- 10.2.2.1
- 10.2.7.6
- 10.2.7.1
- 10.2.7.11
- 10.2.7.12
- 10.2.7.13
- 10.2.7.25

Estos resultados refuerzan la hipótesis de que el nodo 10.2.203.180 es el default gateway de la red. Posiblemente el resto de subgrafos conexos podrían pertenecer al conjunto de máquinas de cada laboratorio. Hemos consultado para confirmar que el default gateway es el que sospechábamos. En este caso, no fue así. El default gateway de la red es el 10.2.203.254. Aunque no se trató del nodo con la menor

cantidad de información de la red, el verdadero default gateway se encontraba dentro de los posibles nodos distinguidos. No hemos podido corroborarlo, pero sospechamos que el nodo 10.2.203.180 puede tratarse del servidor proxy.

Por otro lado, nos comentaron que cada laboratorio tiene un router, y su dirección IP termina con 254. Pudimos observar su presencia en el grafo, pero no aparecieron como nodos distinguidos (excepto el 10.2.203.254).

Hay una serie de nodos distinguidos que no se tratarían de routers. Podrían tratarse de impresoras de red o de algún server. También existe software de seguridad y control, que monitorea la red y podría estar corriendo en estos equipos.

3.3. Red Wifi Wendy's

3.3.1. Cantidad de información dada la fuente binaria S

Este análisis corresponde al experimento realizado sobre la red wifi de un local de Wendy's. A continuación mostraremos los valores calculados y algunas particularidades encontradas en la misma, luego del evaluar los valores obtenidos y tratar de encontrar una respuesta que justifique los mismos.

En primer lugar, tenemos el gráfico correspondiente a la cantida de información según la fuente binaria S, pudiendo observar que la cantidad de información aportada por la aparición de un mensaje unicast es considerablemente mayor que la de un broadcast y de la entropía.

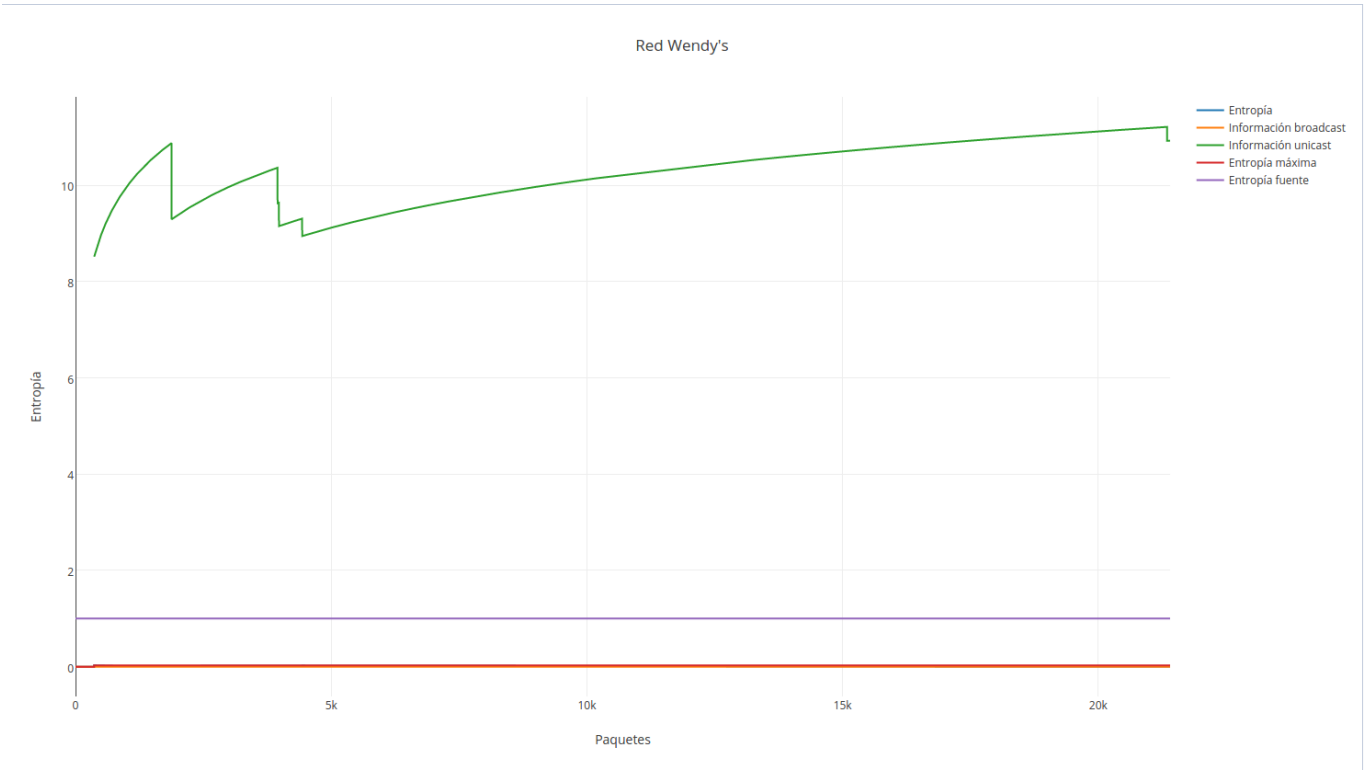

Figura 8: Entropía y cantidad de información de la wifi de Wendy's según la fuente S

Dado que la magnitud de la cantidad de información del unicast es mucho mayor a la del broadcast y la entropía, realizamos un gráfico separado para poder ver con más detalle la evolución de ambas variables.

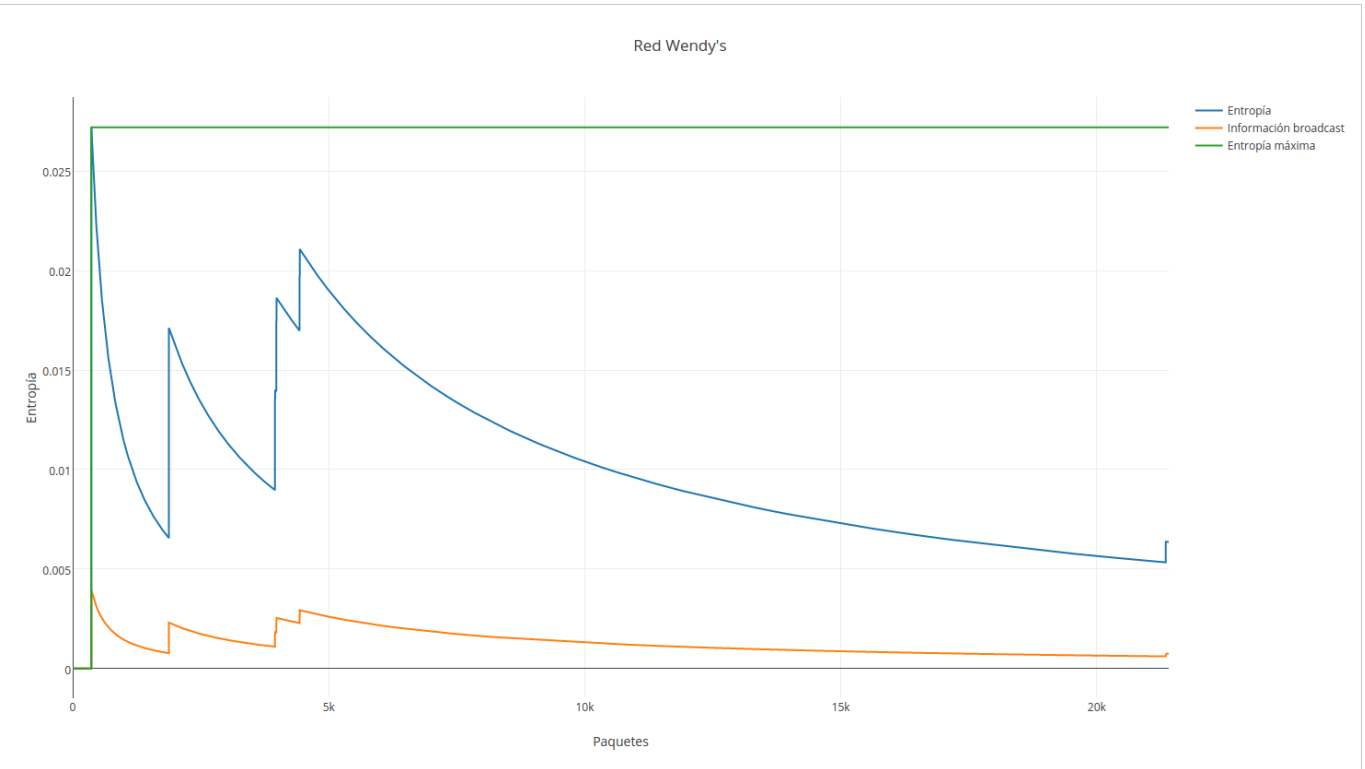

Figura 9: Entropía y cantidad de información del broadcast de la red wifi de Wendy's según la fuente S (zoom)

Podemos concluir, como en el experimento anterior, que la cantidad de información aportada por un unicast es alta debido al gran tráfico de mensajes broadcast en la red con respecto al unicast. Es por ello también que el valor de la entropía es muy baja (cercana a 0), ya que la misma resulta más alta cuando los símbolos de la fuente tienden a la equiprobabilidad. Este valor se mantuvo bajo durante todo el experimento, siendo el máximo valor de entropía alcanzado de 0.02 bits. Éste no era un comportamiento esperado por nosotros. Notamos que hay un gran overhead por parte de los protocolos de control. A continuación, podemos ver algunas características que explican este comportamiento y que nos permiten sospechar que la red no se encontraba funcionando en condiciones ideales.

3.3.2. Red de mensajes ARP subyacente

Ésta se trata de una red mediana, de aproximadamente 50 nodos. Acá comenzamos a notar ciertas anomalías en la red capturada. Notamos que hay una gran cantidad de paquetes ARP cuyo origen y destino coinciden. Podemos observar en la siguiente figura parte de la captura realizada con Wireshark y nos encontramos que la mayoría de los paquetes están etiquetados como *GratuitousARP*. El *gratuitousARP* se trata de un paquete en los que ambas IP de origen y destino coinciden. Este paquete sirve para detectar conflictos de IP en la red. Otra utilidad es la de informar a switches y hosts, de las MACS de sus interfaces para que actualicen sus tablas sin peticiones directas. Más allá de las utilidades, hemos encontrado información que indica que este tipo de paquetes se utilizan para atacar una red, saturando la misma con estos paquetes.

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	a4:6c:2a:85:af:a0	ff:ff:ff:f…	ARP	42	Gratuitous ARP for 172.17.69.137 (Reply)
2	0.000733	a4:6c:2a:85:af:a0	ff:ff:ff:f…	ARP	42	Gratuitous ARP for 172.17.83.172 (Reply)
3	0.001471	a4:6c:2a:85:af:a0	ff:ff:ff:f…	ARP	42	Gratuitous ARP for 172.17.82.91 (Reply)
4	0.106495	a4:6c:2a:85:af:a0	ff:ff:ff:f…	ARP	42	Gratuitous ARP for 172.17.44.197 (Reply)
5	0.110822	a4:6c:2a:85:af:a0	ff:ff:ff:f…	ARP	42	Gratuitous ARP for 172.17.105.120 (Reply)
6	0.208130	a4:6c:2a:85:af:a0	ff:ff:ff:f…	ARP	42	Gratuitous ARP for 172.17.78.226 (Reply)
7	0.208868	a4:6c:2a:85:af:a0	ff:ff:ff:f…	ARP	42	Gratuitous ARP for 172.17.28.165 (Reply)
8	0.313381	a4:6c:2a:85:af:a0	ff:ff:ff:f…	ARP	42	Gratuitous ARP for 172.17.28.190 (Reply)
9	0.524628	a4:6c:2a:85:af:a0	ff:ff:ff:f…	ARP	42	Gratuitous ARP for 172.17.65.60 (Reply)
10	0.625570	a4:6c:2a:85:af:a0	ff:ff:ff:f…	ARP	42	Gratuitous ARP for 172.17.88.149 (Reply)

Figura 10: Wendys - Captura de paquetes en Wireshark

Hemos encontrado una gran cantidad de nodos con diferentes IP a través de la captura de los paquetes ARP. Hemos simplificado el gráfico, mostrando un mínimo porcentaje de nodos. Casi el cien por ciento de los paquetes corresponden a los mencionados anteriormente, en donde origen y destino coinciden.

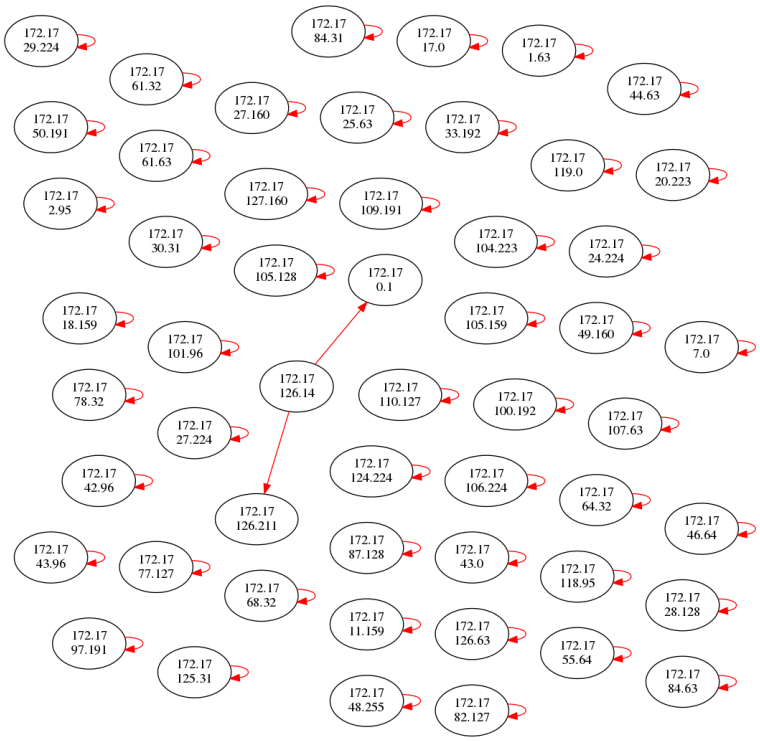

Figura 11: Wendys - Red de mensajes ARP subyacente

Si nos quedamos únicamente con los paquetes cuyo origen difieren del destino, obtenemos el siguiente grafo subyacente:

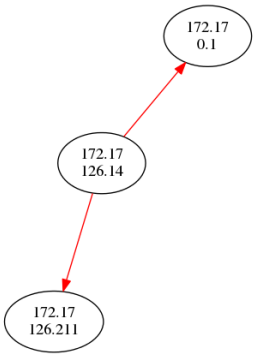

Figura 12: Wendys - Red de mensajes ARP subyacente, considerando paquetes con $src \neq dst$

Únicamente tenemos intercambio de paquetes entre tres nodos: 172.17.126.211, 172.17.126.14 y 172.17.0.1.

3.3.3. Cantidad de información de la fuente S_1

De acuerdo a los cálculos realizados, tenemos que la entropía de la red es 13.81 bits y el máximo es de 14.02. Son valores muy cercanos, lo que indica una distribución bastante uniforme en cuanto a las probabilidades de cada símbolo (no hay un gran conjunto de símbolos con una probabilidad considerablemente mayor a la de los demás).

Debido a la gran cantidad de paquetes y de diferentes IP capturadas, podamos los gráficos para permitir su visualización. Mostrando en la próxima figura las 20 IP con mayor y las 20 IP con menor cantidad de paquetes o cantidad de información.

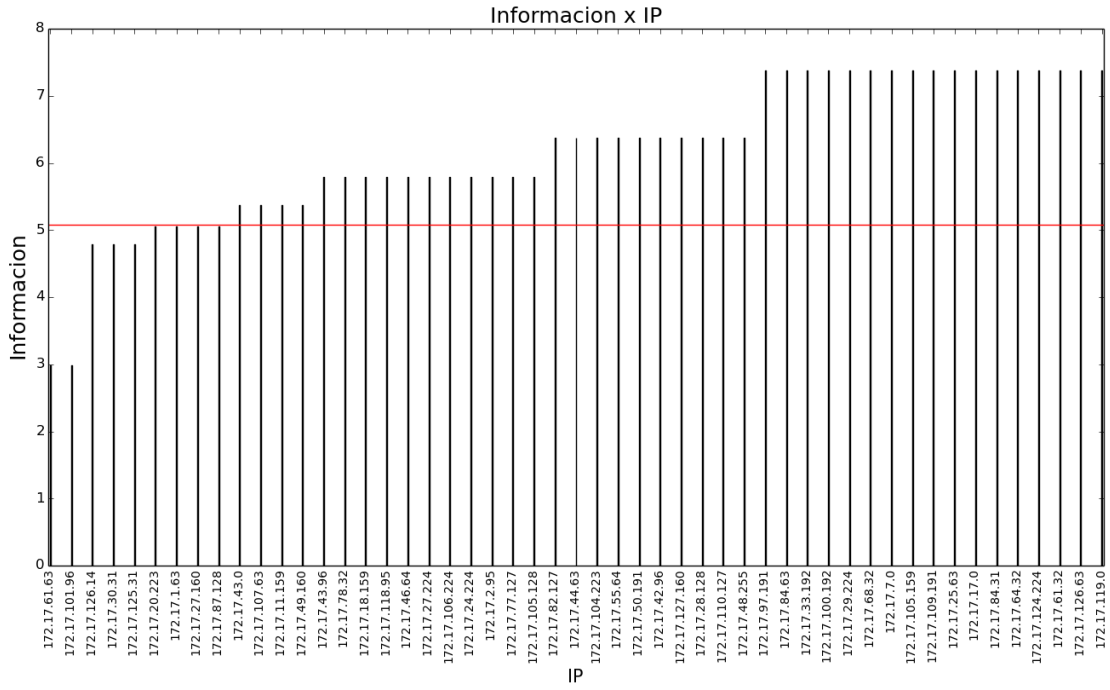

Figura 13: Cantidad de información de la red Wendys de acuerdo a la fuente S_1

Encontramos que el nodo 172.17.82.9 es el que está asociado a una mayor cantidad de paquetes y a una menor cantidad de información, pero no es ninguno de los tres nodos que obtuvimos al quitar los

paquetes con coincidencias entre el destino y el origen.

Si únicamente analizamos la red subyacente correspondiente a los tres nodos, obtenemos el siguiente gráfico:

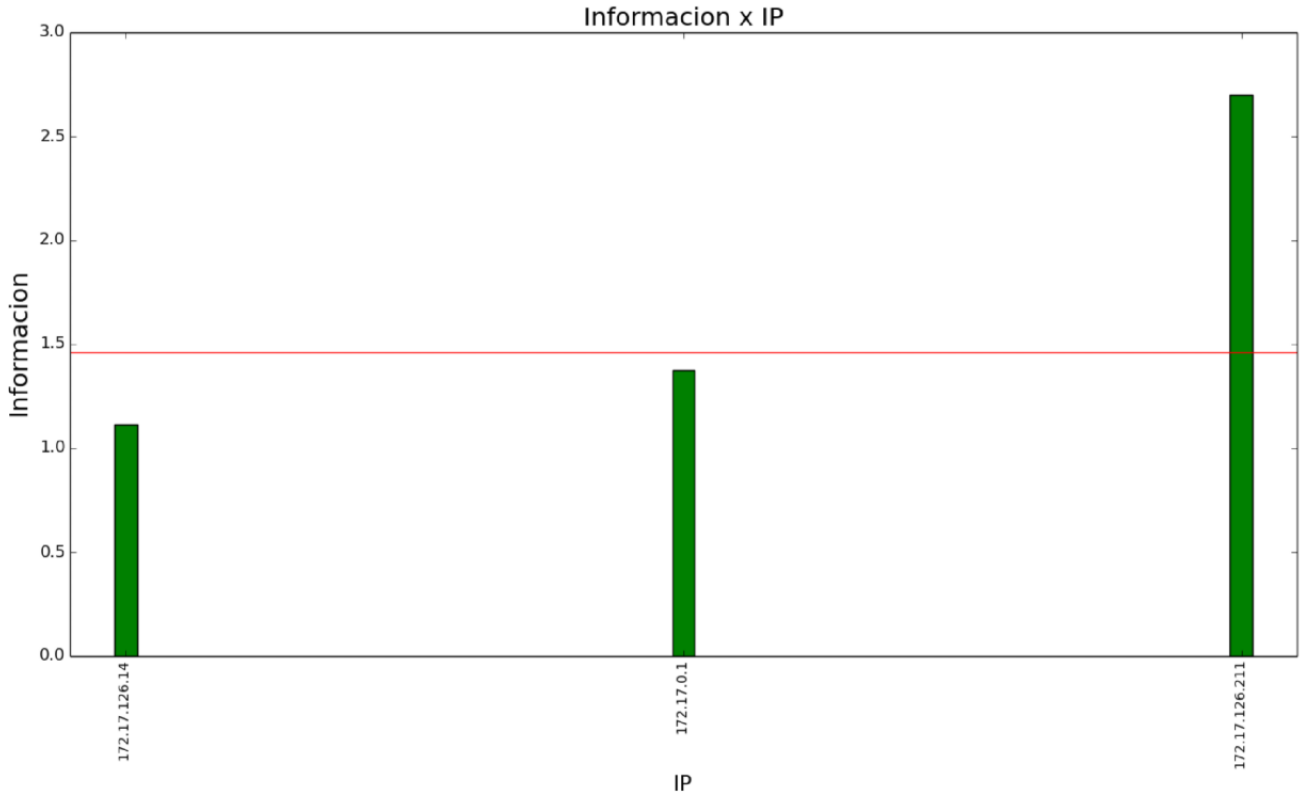

Figura 14: Cantidad de información de la red Wendys de acuerdo a la fuente S_1, considerando paquetes con $src \neq dst$

Como se pudo observar no hubo un gran intercambio de paquetes entre los mismos. No es directa la determinación de un nodo distinguido. En este caso, podemos sospechar que 172.17.0.1 es el $defaultgateway$ más que nada por las características numéricas de la IP.

4. Conclusiones

El presente trabajo nos ha acercado al manejo de herramientas que eran desconocidas hasta el momento para nosotros. La idea de poder escuchar pasivamente los paquetes de una red nos parecía algo mucho más complicado y que requería de ciertas habilidades para vencer controles de seguridad. Sin embargo, nos encontramos con algo que se puede realizar fácilmente sin contar con mucho conocimiento. Estas herramientas de software en combinación con los conceptos de la Teoría de la Información aprendidos nos han permitido identificar nodos distinguidos a través de la valoración de la cantidad de información provista por la aparición de la IP ese nodo dentro de los paquetes ARP (un bajo valor en la cantidad de información del nodo presupone que el mismo es distinguido).
Hemos experimentado con redes de diferentes tecnologías (Ethernet, Wifi). Lamentablemente no pudimos sacar conclusiones al respecto, ya que el tercer experimento no permitió obtener buenos resultados como para realizar una comparación. Capturamos redes de diferentes tamaños. Podemos indicar que la diferencia en la cantidad de información será mayor en redes de mayor tamaño, ya que una mayor cantidad de nodos tratarán de conectarse con el default gateway que en una red de menor tamaño. Además, la entropía será mayor en una red de mayor tamaño, ya que vamos a encontrar una mayor cantidad de nodos con cantidades de información similares y la influencia de la baja cantidad de información provista por el default gateway tendrá menor influencia que en una red de menor tamaño. La entropía tiene gran importancia en la detección de default gateways. Si se trata de una red básica (esquema tradicional de múltiples equipos conectados a routers), seguramente podremos detectarlos al analizar la cantidad de información provista en esos nodos, quedándonos con aquellos que proveen la menor cantidad de información. Cuanto mayor sea la cantidad de nodos, mayor será la entropía, como explicamos líneas arriba. Esto posibilita también que aparezca una mayor cantidad de nodos distinguidos en la red (cualquier nodo que intercambie un número no muy considerable de paquetes ARP mayor a la media, resultará con una cantidad de información menor a la entropía).