TP1: Wiretapping

Matías Millassón
Lautaro Leonel Alvarez

FCEN - Universidad de Buenos Aires

11 de julio de 2017

Resumen

En este trabajo se experimentó con cuatro redes de diversos tamaños, en particular dos hogareñas y dos de lugares bastantes concurridos. Se utilizó un sniffer para experimentar con la red y para conocer su topología entre otras cosas.

I. Introducción

En el siguiente trabajo práctico vamos a análizar 3 redes locales con el fin de caracterizarlas en base a los nodos que participan enviando o recibiendo paquetes, la cantidad de paquetes y el rol que cumplen (o creemos que cumplen) dentro de la red. Para esto, utilizaremos un *sniffer* que escucha la red y almacena los datos recolectados.

Antes que nada, daremos unos conceptos importantes sobre el tema para poder explicar mejor el funcionamiento y alcance de esta herramienta.

i. ARP: Adress Resolution Protocol

Es el protocolo encargado de mapear direcciones del nivel 3 al nivel 2. Los paquetes ARP poseen varios campos de información pero en este trabajo solo nos van a interesar la dirección MAC e IP destino y fuente y el tipo de mensaje, es decir si es who-has o is-at.

Cuando un equipo quiere comunicarse con otro envía un paquete ARP de tipo who-has (broadcast) y el equipo buscado responde con un paquete ARP de tipo is-at (unicast). Estos envíos sirven para que cada equipo arme su propia tabla ARP en la cual asocia direcciones IP con direcciones MAC.

ii. Modo promiscuo

Se trata de una configuración de la placa de red, en la cual transmite a sus niveles superiores todos los paquetes que escucha, y no solo los que van dirigidos a nuestra pc (como pasa normalmente). La herramienta sniffer utiliza precisamente este modo para tomar todos los paquetes que logre leer.

iii. Teoría de la información

Vamos a usar la teoría estadística de la información (o teoría de Shannon). Sin entrar tan en detalle, nos interesarán las siguientes definiciones:

- **Información de un suceso:** Dado un suceso i de una fuente cuya probabilidad es estrictamente mayor que cero se define su información como $c_i = -log(p_i)$ con p_i la probabilidad de que ocurra el suceso i. Se puede observar que cuanto menos probable sea un suceso más información provee y cuanto más probable sea menos información aporta.
- **Entropía de una fuente:** Dada una fuente S se define la entropía como $H(S) = \sum_{i=1}^{\#(S)} p_i * c_i$ con c_i la información del suceso i y p_i su probabilidad. Esta métrica es un promedio ponderado de la información que brinda cada suceso.

IV. Herramienta sniffer

Ahora sí, podemos pasar a definir la herramienta que desarrollamos y una breve introducción a su funcionamiento. La herramienta utiliza una librería llamada scapy[1], que es la encargada de pasar al sistema a modo promiscuo y obtener todos los paquetes que pueda identificar. De esta manera, luego de obtener estos paquetes, se calculan la información y entropía de distintas fuentes de información (que luego mencionaremos) y al mismo tiempo se almacena la información que consideramos relevante en un archivo de salida (para poder ser utilizado luego).

V. Fuentes de información analizadas

Para analizar los datos de paquetes obtenidos por la herramienta definimos dos fuentes de información:

- **S:** El conjunto de símbolos definidos es $\{S_{broadcast}, S_{Unicast}\}$. El símbolo $S_{broadcast}$ corresponde a los paquetes capturados cuya dirección de destino es ff:ff:ff:ff:ff:ff, mientras que el resto de los paquetes van a ser considerados $S_{unicast}$.
- **S1:** El conjunto de símbolos definidos son todas las direcciones IPv4 que tuvieron todos los dispositivos mientras estaban conectados a la red. En esta fuente sólo consideramos los paquetes de tipo **who-has**.

II. Primera captura: Laboratorio de Pabellón 1

Para una primera experimentación, decidimos correr la herramienta sniffer en la red interna del pabellón 1 de Ciudad Universitaria. Nos conectamos por Wifi a la red y tomamos muestras durante 17 minutos. Utilizamos la fuente de información S1 para diferenciar cada host como un símbolo. Luego, calculamos la información de cada host y con esto, la entropía de la fuente (red).

[1]SCAPY: https://github.com/secdev/scapy/

I. Resultados y análisis

En el gráfico de la figura 1 se pueden ver los 25 hosts con menor información. La linea punteada roja representa la entropía de la fuente y, como se observa rápidamente, hay un solo host que se encuentra del lado izquierdo (**10.210.210.199**). Esto significa que es un símbolo con muy poca información y por ende, mucha probabilidad. Para nosotros, esto significa que es un host que participa mucho de la red, preguntando constantemente qué dirección MAC corresponde a una dirección IP. Tomamos entonces a este host como un nodo distinguido de la red.

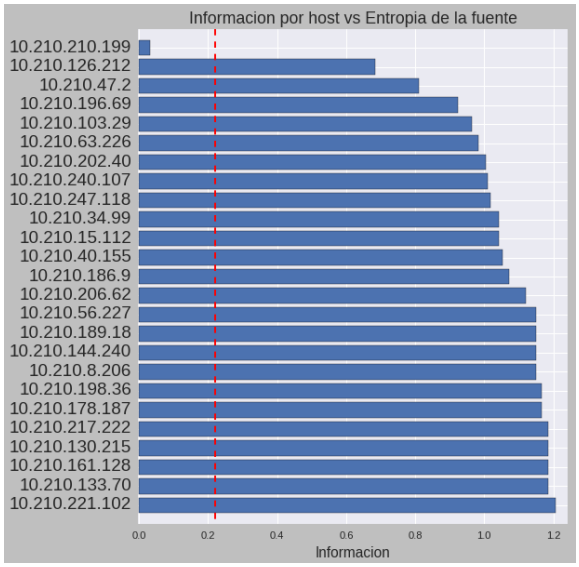

Figura 1: *Información de cada símbolo (host) de la red comparada con el valor de la entropía de la fuente de información (red local). Se limita el gráfico a los 25 símbolos con menor valor de información.*

La distinción del host con IP **10.210.210.199** la podemos corroborar en el gráfico de la figura 2. Este grafo representa a la red relacionando nodos con hosts y aristas con mensajes. El diámetro de los nodos está dado por el nivel de participación que tiene en la captura tomada. Asimismo, se obviaron muchos nodos con comportamiento similar que no sumaban valor al grafo, asegurando que los mas *participativos* se mantengan. Este valor de *participación* fue

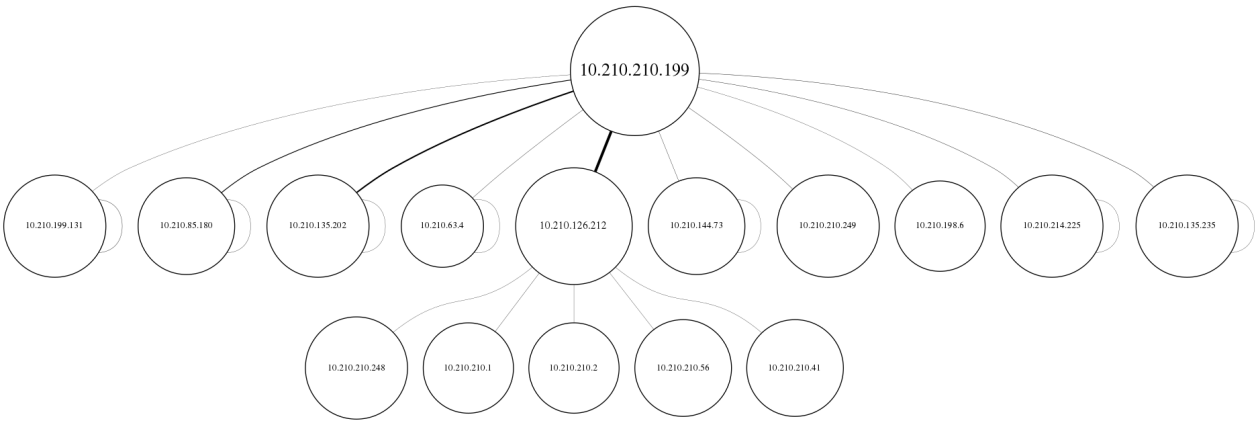

Figura 2: *Grafo con los nodos mas interesantes de la red de la primer captura. El diámetro del nodo implica mayor participación.*

calculado teniendo en cuenta los siguientes parámetros:

1. Cantidad de mensajes ARP enviados (solo tomando los del tipo who-has).
2. Cantidad de mensajes ARP que consultaban por su ip (solo tomando los del tipo who-has).
3. Cantidad de hosts *distintos* por los cuales preguntó su dirección MAC.
4. Cantidad de hosts *distintos* que preguntaron por su dirección MAC.

A estos 4 parámetros se les asigna un peso, que depende de lo que queramos observar. En este caso en particular, los parámetros 1 y 2 tienen peso 0 porque queríamos enfocarnos en los 3 y 4, que en cierta forma implican a los 1 y 2 (si la cantidad de *distintos* es grande, esto implica que la cantidad sin tener en cuenta distinción de hosts también es grande). Al parámetro 3 le asignamos peso 2, mientras que al 4 le asignamos peso 10. De esta manera, podemos destacar fuertemente los hosts por los cuales muchos otros consultaron por él, pero también teniendo en cuenta (con menor importancia) los hosts que consultaron por muchas direcciones. Lo que logramos con estos valores, es lo que observamos en el grafo de la figura 2.

Las aristas por su parte, representan la consulta de un host por la dirección MAC de otro (sin diferenciar la dirección de los mensajes). Y el peso de cada arista (relacionado con el grosor) está dado por la cantidad de mensajes que relacionan a los hosts.

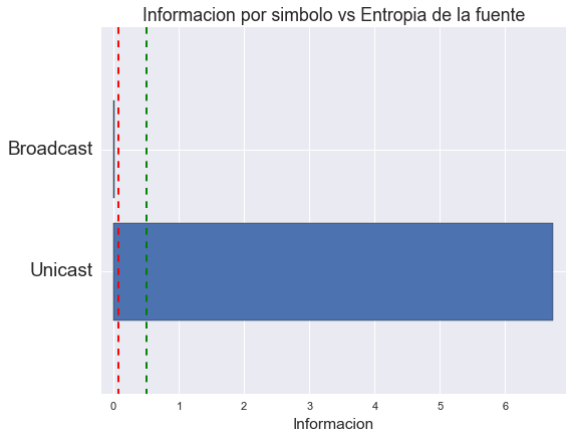

Figura 3: *Información de cada símbolo de la fuente S comparado con el valor de la entropía.*

Observando este grafo, notamos que el host con IP **10.210.126.212** tiene un comportamiento particular. Para analizar mejor la situación vamos a referirnos al host con IP **10.210.210.199** como **Nodo A** y al host con IP **10.210.126.212** como **Nodo B**. A continuación vamos a mencionar información que extrajimos de la muestra y luego pasaremos a analizar el por qué de estos valores y tratar de caracterizar los nodos según su comportamiento:

- Solo el nodo A preguntó por la dirección del nodo B, y lo hizo 121 veces.
- El nodo B nunca preguntó por la dirección del nodo A, mientras que otros 268 hosts si lo hicieron.
- El nodo A realizó 5796 preguntas, por 298 direcciones distintas (entre ellas la ip del nodo B).
- El nodo B realizó 151 preguntas, por 101 direcciones distintas.

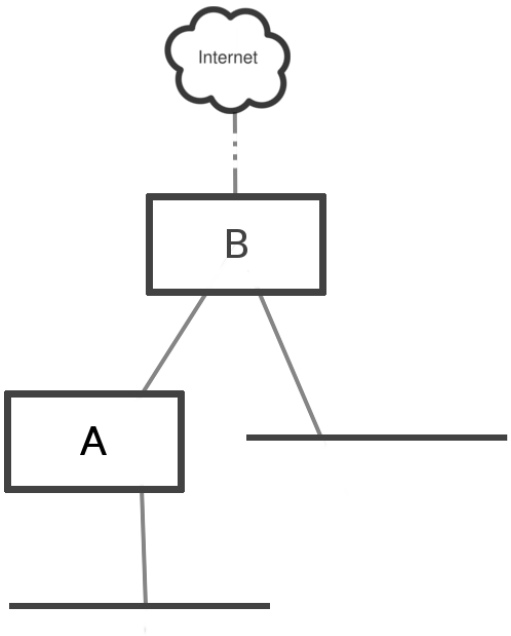

Figura 4: *Diagrama de cómo creemos que está organizada la red de la primera captura.*

II. Conclusiones

Por esta información, creemos que el nodo A es un access point de la red Wifi, ya que tiene interacción con muchos hosts (entre 268 y 297) y la gran mayoría de ellos solo preguntan por él. Y en el caso del nodo B, creemos que es algún router por el cual debe acceder el nodo A para poder salir a internet. Al mismo tiempo, el nodo B tiene interacción con bastantes hosts (151). También vemos que no tienen interacciones en común (quienes preguntan por A no preguntan por B, y la inversa). Por esto, creemos que el nodo B debe ser el gateway de otra red local. En la figura 4 planteamos la idea de lo que creemos que puede ser la red, dejando varios interrogantes, pero dando una idea general de su organización.

En la figura 3 podemos ver la diferencia entre las informaciones del símbolo *Broadcast* y el símobolo *Unicast*. En rojo observamos el valor de la entropía de la fuente S, mientras que en verde podemos ver la entropía máxima (correspondiente a 1/2). Notamos que la entropía de la red está por debajo de la máxima debido a que la cantidad de mensajes Broadcast superó ampliamente a la cantidad de mensajes Unicast de la red.

III. Segunda Captura: Red local chica

Para una segunda experimentación, tomamos una captura de una red casera por medio de la herramienta sniffer antes mencionada. Se capturaron mensajes durante 2 horas y al ser una red pequeña, a diferencia del primer experimento, vamos a analizar tanto la fuente de información S como la S1.

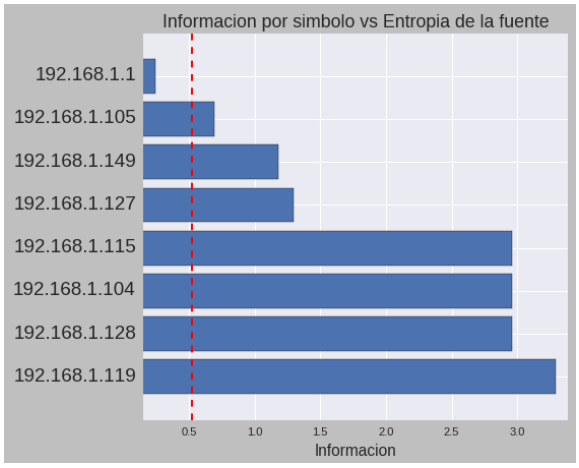

Figura 5: *Información de cada símbolo (host) comparada con el valor de la entropía de la fuente de información (red local).*

I. Resultados y análisis

En el gráfico de la figura 5 podemos observar la información de cada símbolo (host) en la fuente S1. Se puede ver claramente que hay un solo host que se encuentra a la izquierda de la linea punteada roja, que representa la entropía de la red. Este host con IP **198.168.1.1** va a ser para nosotros un nodo distinguido, ya que participa fuértemente en los mensajes ARP que se envían en la red y por esto, tiene un valor de información pequeño.

Con respecto a la fuente de información S, tenemos el gráfico de la figura 6. En este gráfico observamos la información del símbolo *Unicast* y el símbolo *Broadcast*. Podemos notar que, al igual que en la primer captura, el símbolo *Unicast* supera ampliamente al símbolo *Broadcast*, lo que implica que se identificaron una cantidad elevada de mensajes del tipo *Broadcast* en la red. Pero en contraste con el caso anterior, la diferencia no es tan amplia.

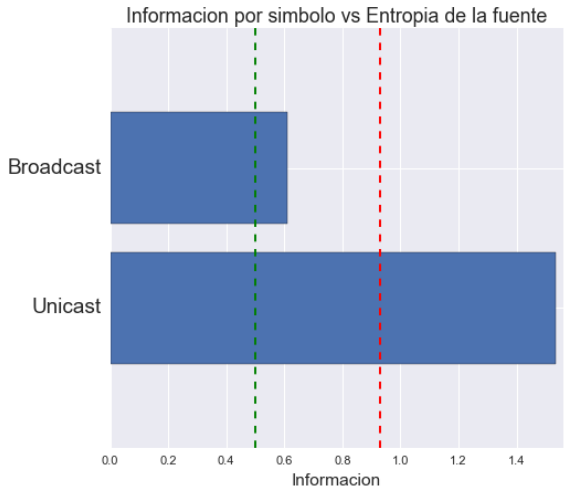

Figura 6: *Información de cada símbolo (Broadcast / Unicast) comparada con el valor de la entropía de la fuente de información (red local).*

El grafo de la figura 7 es una representación de la red de acuerdo a los hosts identificados y las relaciones entre ellos (dadas por mensajes ARP que consultan direcciones entre ellos). El tamaño de los nodos, al igual que en la primer captura, está relacionado con la *participación* que tienen en la red. Aquí también podemos notar que el host con IP **192.168.1.1** es el que mas participa de los mensajes ARP. También podemos notar que todos los otros hosts se encuentran unidos a él (preguntaron por su dirección o la inversa). De el mismo grafo, podemos ver también que los hosts **192.168.1.104** y **192.168.1.129** se encuentran conectados, mientras que todo el resto solo se conectan con el host **192.168.1.1**. Luego de corroborar los datos, notamos que hay un mensaje ARP que emitió el host **192.168.1.104** preguntando por la dirección del host **192.168.1.129**.

II. Conclusiones

Como observamos en el grafo y corroboramos con la información del símbolo en la fuente de información S, el host **192.168.1.1** se encuentra conectado con todos los otros hosts, por lo que creemos que es un router o access point con el cual el resto de los hosts se comunican. El resto de los hosts, nos resultan similares y no creemos que sean hosts destacados o distinguidos.

IV. Tercer captura: Red local muy chica

Se realizó una tercer captura en una red local que poseía solamente tres hosts. La captura duró aproximadamente veinte minutos. En este caso vamos a analizar ambas fuentes de información.

En este experimento se tuvo una red con tres hosts, uno es el router (**192.168.1.1**), otro es dónde se estaba ejecutando la herramienta de monitoreo (**192.168.1.124**) mientras que en el equipo restante se estuvo realizando una descarga mediante un cliente torrent durante casi todo el tiempo (**192.168.1.129**).

I. Resultados y análisis

En el gráfico de barras en el que se compara la información de cada símbolo la fuente S con la entropía de la misma (figura 8) se puede ver que, al igual que en los demás experimentos

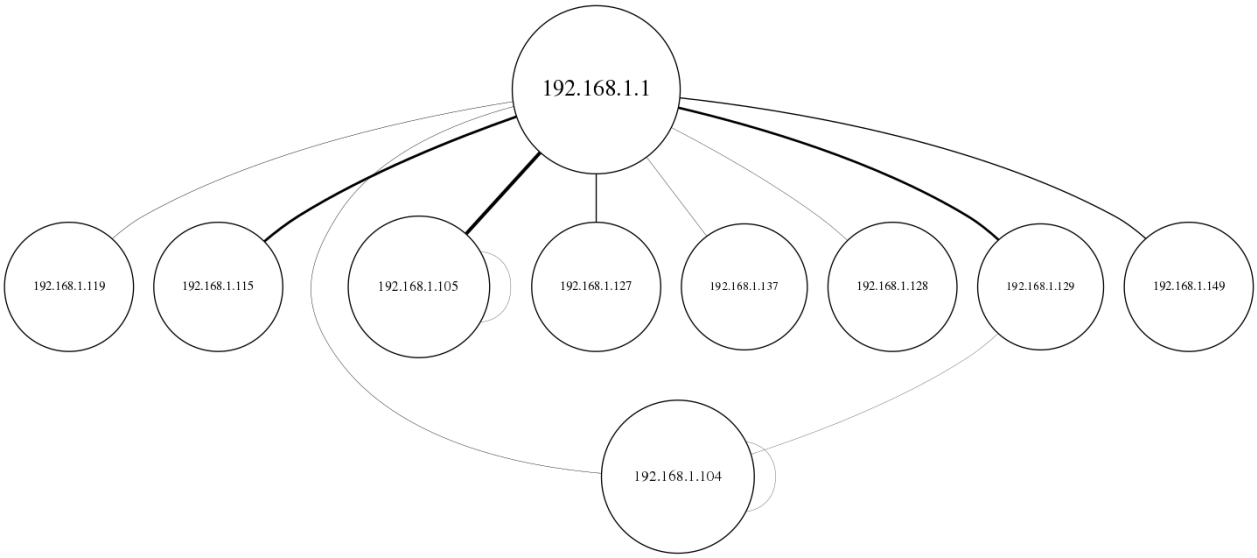

Figura 7: *Grafo con los nodos de la red de la segunda captura. El diámetro del nodo implica mayor participación en el envío de paquetes.*

hubo muchísimos más paquetes *Broadcast* que *Unicast*. Aún así el valor de la entropía está bastante cerca de ser el máximo, el cual es 0.5 ya que son dos símbolos.

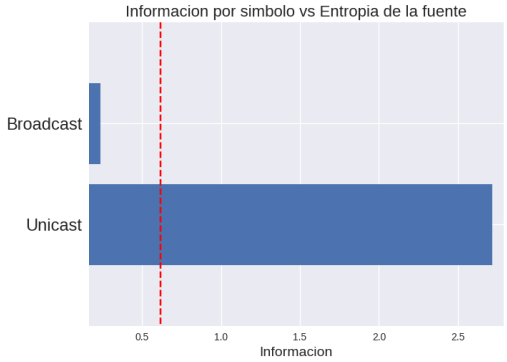

Figura 8: *Información de cada símbolo (Broadcast / Unicast) comparada con el valor de la entropía de la fuente de información (red local).*

Con respecto al gráfico de barra de los hosts (figura 9), el equipo en el que se estuvo realizando la descarga tuvo una información muy baja y fue el único que estuvo por debajo de la entropía. Este resultado muestra una diferencia con las demás redes ya que el default gateway no resultó ser el de mayor probabilidad. Pero, es importante comentar que el host en el que se estaba iniciando una descarga envió paquetes ARP consultando por todas las direcciones posibles de la red. Por simplicidad del grafo no los incluimos a todas las direcciones de la red. También, en el grafo se puede ver una fuerte relación entre el host de la herramienta de monitoreo con el router como así también la de éste último con el equipo donde se estaba realizando la descarga.

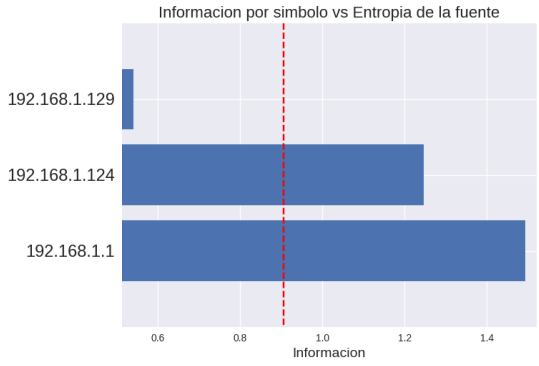

Figura 9: *Información de cada símbolo (host) comparada con el valor de la entropía de la fuente de información (red local).*

II. Conclusiones

La entropía máxima es 0.5 ya que son 2 símbolos pero en este caso no se logró llegar a dicho valor. Esto sugiere que hay muchos envíos de mensajes de control, secuencia, etc. ya que éstos suelen ser broadcast. Este resultado está muy relacionado con el overhead de la red ya que los mensajes de control ocupan ancho de banda que, en algún caso, podrían enviarse datos.

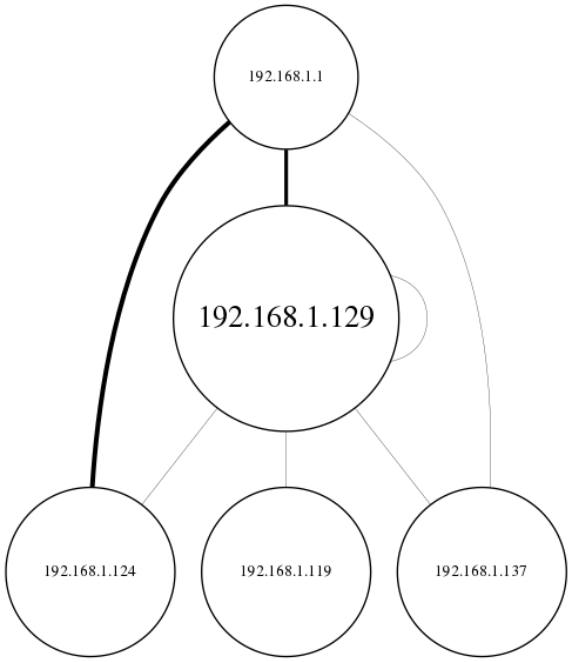

Figura 10: *Diagrama de cómo creemos que está organizada la red de la tercer captura.*

En esta red se puede distinguir solamente un nodo (**192.168.1.129**) pero fue un comportamiento por demás anómalo, por lo que no le podemos asignar ninguna función en especial dentro de la red. Creemos que si la captura hubiera durado alrededor de dos horas, el router (**192.168.124.1**) hubiera tenido mucha más participación. En particular por ser una red muy pequeña es esperable que se distinga un nodo, aunque si es inesperado qué nodo resultó ser.

Lamentablemente no hay una correspondencia entre lo que se conoce de la red y el resultado del experimento ya que la captura no duró lo suficiente y este outlier no nos permitió dilucidar la identidad de la puerta de enlace predeterminada.

V. Cuarta captura: red intermedia

En este último experimento, ejecutamos nuestro sniffer en una casa de comidas rápidas muy concurrida. La red es más grande que una hogareña pero más pequeña que la presentada en el primer experimento. La captura duró aproximadamente media hora.

I. Resultados y análisis

Como esta red posee muchísimos nodos decidimos realizar el gráfico con los que tienen menor información, en el mismo (figura 11) se pueden ver que hay algunos hosts con menos información que la entropía, mientras que hay otros que no. Pero en realidad todo los hosts restantes, es decir los que no fueron incluidos en el gráfico, tienen una información mayor. El host de la ip terminada en 110 es la puerta de enlace predeterminada, por lo cual es lógico que tenga mucha interacción con otros hosts. En la red hubo en total 127 hosts.

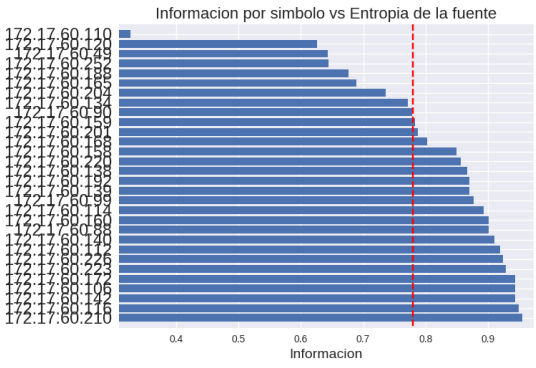

Figura 11: *Información de cada símbolo (host) comparada con el valor de la entropía de la fuente de información (red local).*

Por otra parte, en el gráfico de barras en el que se compara la información de los paquetes *Broadcast* contra la de los paquetes *Unicast*

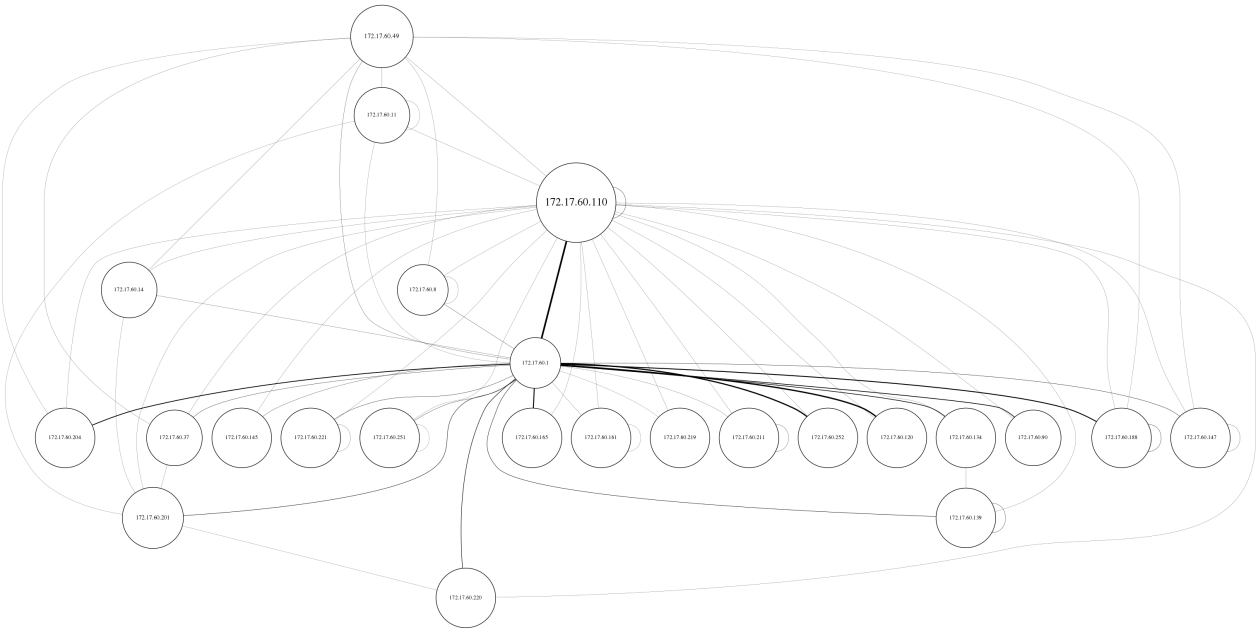

Figura 12: *Diagrama de cómo creemos que está organizada la red de la cuarta captura.*

(figura 12) se puede ver que hubo una gran diferencia a favor de la cantidad de paquetes broadcast en cuanto a paquetes enviados. Esto es sumamente razonable ya que la gran mayoría de los dispositivos conectados a la red son clientes.

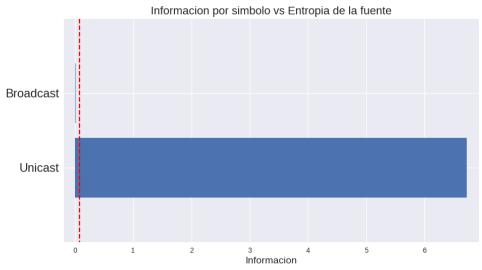

Figura 13: *Información de cada símbolo (Broadcast / Unicast) comparada con el valor de la entropía de la fuente de información (red local).*

Con respecto al grafo en el que mostramos los nodos distinguidos, podemos ver que el nodo **172.17.60.110** tiene muchísima más participación que el resto de los nodos. En particular, notamos que preguntó aproximadamente una vez por cada nodo pero, en cambio, sólo recibió preguntas de él mismo y del host **172.17.60.49**. Además, el nodo correspondiente al equipo cuya IP es **172.17.60.1** suele ser muy solicitado por la gran mayoría de los equipos de la red, pero en contrapartida éste no consultó en ningún momento por los demás.

II. Conclusiones

En este caso, la fuente S no tiene entropía máxima ya que hay muchos mensajes de control (por la gran cantidad de nodos activos que tiene la red). Además, por la longitud de la captura se acentuó esta tendencia, ya que muchos equipos no permanecieron durante toda la captura (recordemos que es un local de comida rápida). De esta manera hay mucho overhead en esta red.

En esta red se pudieron distinguir varios nodos (nueve en total). Son bastantes pocos con respecto al tamaño de la red pero, aún así, es razonable porque una red que tenga todos los nodos con un tráfico parejo no es conveniente. Pudimos notar que hubo un nodo que se comportó de manera no esperada, el nodo **172.17.60.1**. Este nodo hacía de intermediario

entre la gran mayoría de los nodos de la red y el default gateway (**172.17.60.110**).

Gracias al resultado de la captura podemos afirmar que el default gateway es el equipo cuya ip es **172.17.60.110**. Nos convencemos también de que la herramienta también sirve para detectar puntos intermedios de paso hasta llegar al router.

VI. Conclusiones generales

En este informe realizamos un recorrido por cuatro tipos de redes locales con características y dimensiones distintas. La idea inicial era que las redes fueran bastante disímiles para poder contrastarlas e identificar rasgos que nos resulten llamativos. Nos basamos principalmente en el tamaño de la red y no en la tecnología de los equipos que se encontraban en ellas. Sin embargo, al trabajar sobre redes grandes notamos algunos factores que nos daban información sobre la organización de dichas redes y, luego de analizar esto, dimos una opinión sobre los posibles equipos que participan en ella.

Luego de observar el comportamiento de las redes con mayor cantidad de nodos, notamos que los mensajes Broadcast superan ampliamente a los mensajes Unicast. Esto se lo debemos, como es de esperarse, a los mensajes de control que se necesitan para mantener la red funcionando correctamente y pudimos notar el impacto directo que tiene en la información de dicho símbolo (*Broadcast*). En contrapartida, la información de los mensajes Unicast es elevada, y todo esto hace que la entropía diste mucho mas de su valor máximo (*0.5*).

Para el caso de la fuente S1, notamos que cuanto mas grande es la red, mas notamos la distinción de un nodo router o access point (que en general es planteado como el gran candidato a ser un nodo distinguido). De la misma manera, se observaron algunos otros nodos que se acercaban (que luego intentamos identificar como .ºrganizadores"de la red) y muchos nodos que se comportan de manera similar (los host consumidores de la red).

Sobre la misma idea de distinguir nodos, este trabajo nos deja la idea de que la comparación de la información de los distintos símbolos (hosts) con la entropía de la fuente S1 puede brindarnos información relevante sobre los nodos distinguidos de la red. Y al seguir por ese camino, cuando ordenamos los hosts por su información notamos que varios, a pesar de posicionarse por encima de la entropía, también brindaban información sobre la organización de la red.