Teoría de las Comunicaciones

Segundo cuatrimestre 2017

Departamento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Informe

TP1

Integrante	LU	Correo electrónico
Alejandro Ferrante	371/09	
Gonzalo Guillamon	97/12	
Malena Ivnisky	421/12	malenaivnisky@gmail.com

Índice

1. Introducción

En este trabajo práctico nos proponemos expermientar con el trafico de redes a traves de la tecnica de wiretapping. Para ello monitorearemos tres redes distintas: dos redes domésticas privadas y una red abierta pública e inalambrica. Las analizaremos en función de dos modelos de fuentes diferentes para poder asi comparar los resultados en funcion de la teoria de la Informacion y ver las diferencias en el trafico de paquetes de cada una teniendo en cuenta las particularidades que presentan tanto redes abiertas y cerradas asi como inalambricas o fijas. El objetivo ultimo del presente trabajo es corroborar la utilidad de la teoria de la informacion como herramienta de analisis par encontrar hosts o protocolos distinguidos.

2. Justificación de la elección de la fuente S2

La fuente S2 fue diseñada para poder distinguir a los hosts de cada red, usando las IP de los paquetes ARP capturados. Propusimos un modelo en el cual los símbolos estan conformados por las IPs de los emisores de paquetes ARP del tipo who-has. Esto permitira distinguir los distintos actores dentro de la red ya que cada uno debera en algun punto mandar este tipo de paquetes para poder llenar sus tablas MAC.

3. Experimento 1: Red de Starbucks

3.1. Descripción del contexto

El primer experimento fue realizado en una red inalambrica abierta de FibertelZone, por medio de una conexión Wi-Fi en una sucursal de la cadena Starbucks (ubicada en Avenida Corrientes esquina Malabia, en la ciudad autonoma de Buenos Aires, Republica Argentina) el dia Lunes 9 de Octubre de 2017. Al momento de comenzar las mediciones, se observo que dentro de la red se encontraban conectadas 10 notebooks y varios dispositivos mobiles (se estima uno por persona) , además del router propio del local. Considerando aproximadadmente 30 personas en el lugar contamos con 41 dispositivos conectados.

3.2. Descripción de la captura

La captura resulto en un total de 12000 paquetes. La figura 1 muestar la distribucion de los protocolos de los paquetes capturados. Se puede ver claramente que la mayoria de los paquetes son de tipo Ipv4, mientras que en menor medida encontramos paquetes de tipo Ipv6 (0x86dd) y un menor porcentaje de paquetes ARP. Los dos primeros son distintas versiones del protocolo Ip (nivel de redes dentro del modelo OSI), mientras que el ultimo es un protocolo utilizado por los dispositivos para llenar su tabla de drecciones MAC, ya que cada dispositivo necesita para mandar correctamente un paquete cual es la direccion local (MAC) correspondiente a una direccion Ip. Resulta esperable que la mayoria de los paquetes sean del protocolo Ip, ya que es el utilizado para el intercambio de datos. Sin embargo, consideramos tambien (comparativamente) relevante la cantidad de paquetes ARP, lo cual atribuimos a la gran cantidad de dispositivos conectados en la red analizada.

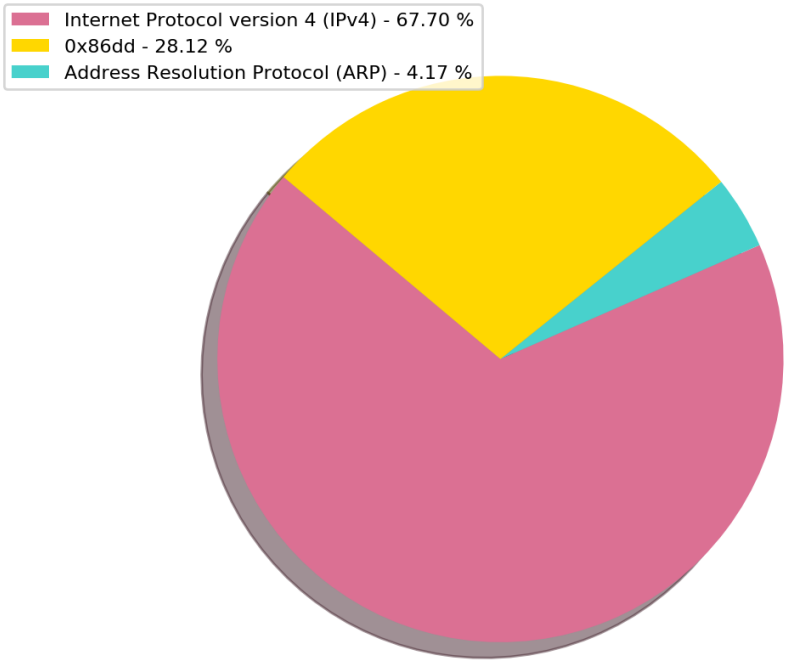

Figura 1: Gráfico que muestra la distribución de protocolos en la red.

En la figura 2 comparamos el porcentaje de paquetes broadcast contra el porcentaje de paquetes unicast. Un paquete bradcast es aquel cuya direccion destino tiene un valor especial, el cual al ser recibido siempre se considera correspondiente (a diferencia de un paquete unicast donde sera considerado correspondiente cuando la direccion destino coincida con la propia). En este caso observamos que los paquetes broadcast representan el 16Como muestra la figura 3, los protocolos con mensajes broadcast son Ipv4 y ARP, siendo estos ultimos del tipo who-has. Este tipo de mensaje corresponde a un pedido dirigido a todos en la red solicitando la direccion MAC de una Ip determinada.

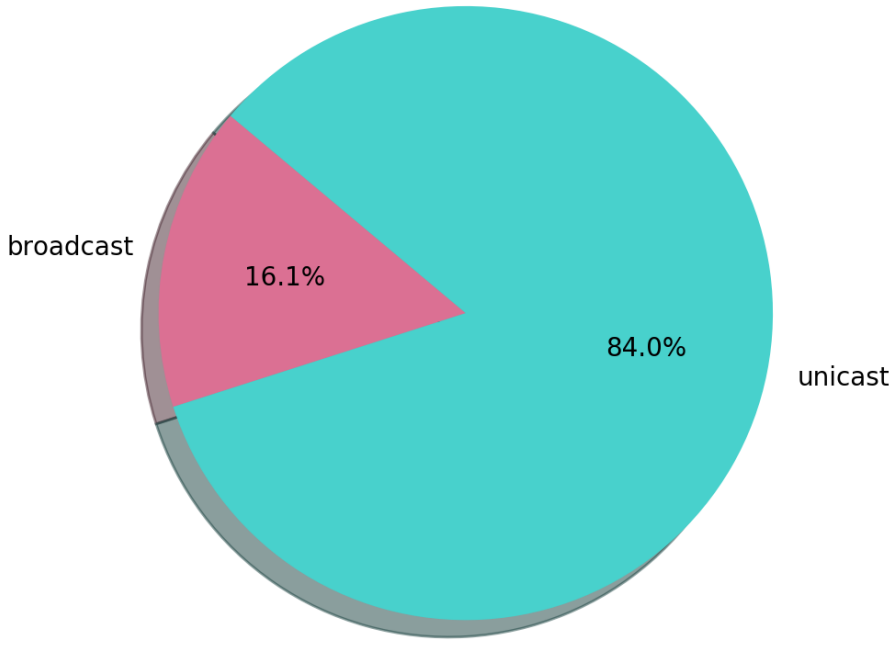

Figura 2: Gráfico que muestra los porcentajes de tráfico broadcast y unicast.

3.3. Análisis de la captura

El modelado de la fuente S1 fue realizado siguiendo las especificaciones de la consigna. En la figura 3 se puede observar que separamos a los paquetes por su protocolo y por si son broadcast o unicast y calculamos su informacion. Incluimos tambien la entropia maxima y la entropia muestral. La probabilidad de aparicion de un simbolo es inversamente proporcional a su informacion. Para esta captura particular el símbolo con más información es representado por los paquetes unicast de tipo ARP. Estos paquetes son los is-at (las respuestas a los who-has).

El símbolo con más información (ARP unicast) tiene aproximadamente 10 veces la información del menor (IPv4 unicast). Notese que segun la teoria de la informacion, la entropía máxima se alcanza cuando la información de todos los símbolos es la misma.

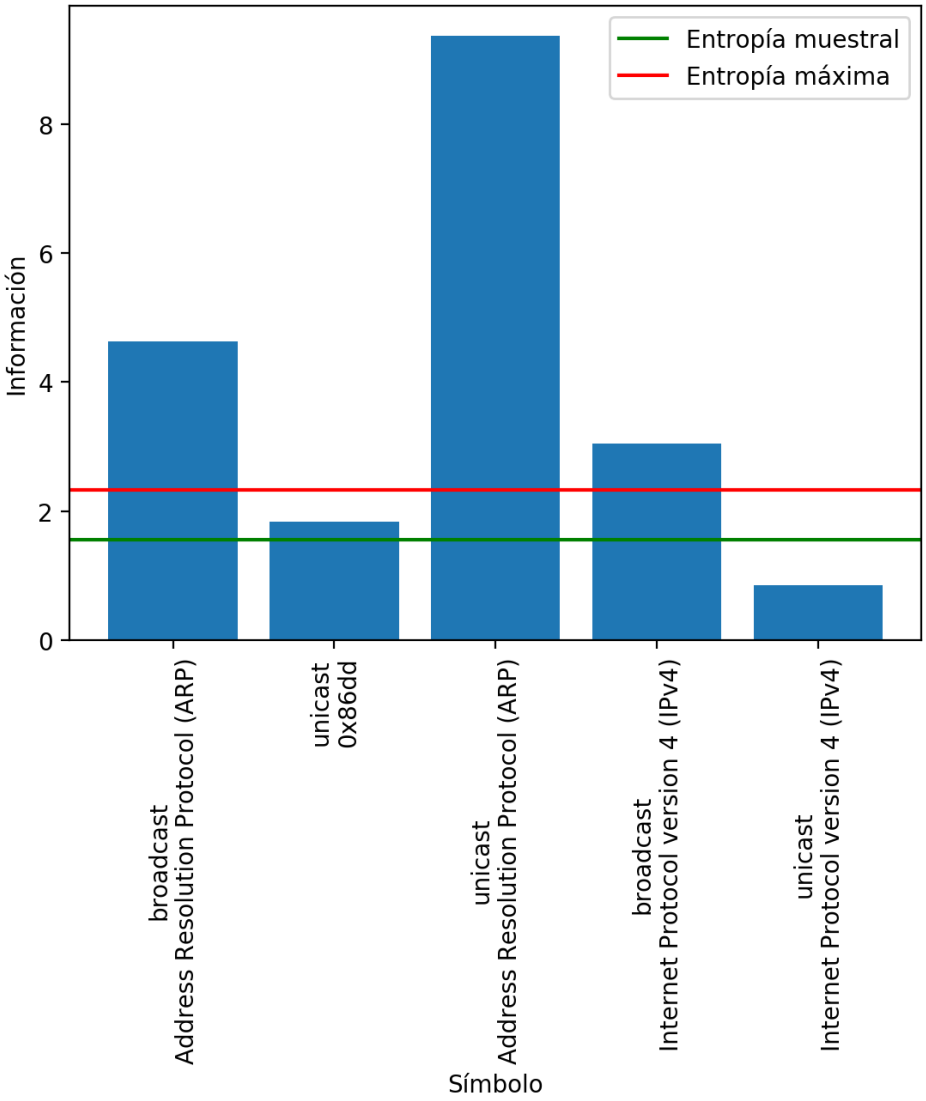

Figura 3: Gráfico de la información de los símbolos de la fuente S_1 observados en esta red. Se muestra la entropía muestral de S_1 y su entropía máxima.

El análisis usando el modelado de la fuente S2 (explicado anteriormente) se puede ver en la figura 4. En este caso notamos que los símbolos de la fuente podrían dividirse en 2 o 3 grupos de acuerdo a su información. Se observan algunos puntos distinguidos, en particular las 8 IPs con información menor a la entropía de S2. Estas son las IPs que más requests ARP hicieron.

La entropía muestral es aproximadamente 4/5 de la máxima. Encontramos que este nivel se debe a la presencia de varios símbolos con valores de información similares (nuevamente, la máxima entropia se alcanza con equiprobabilidad).

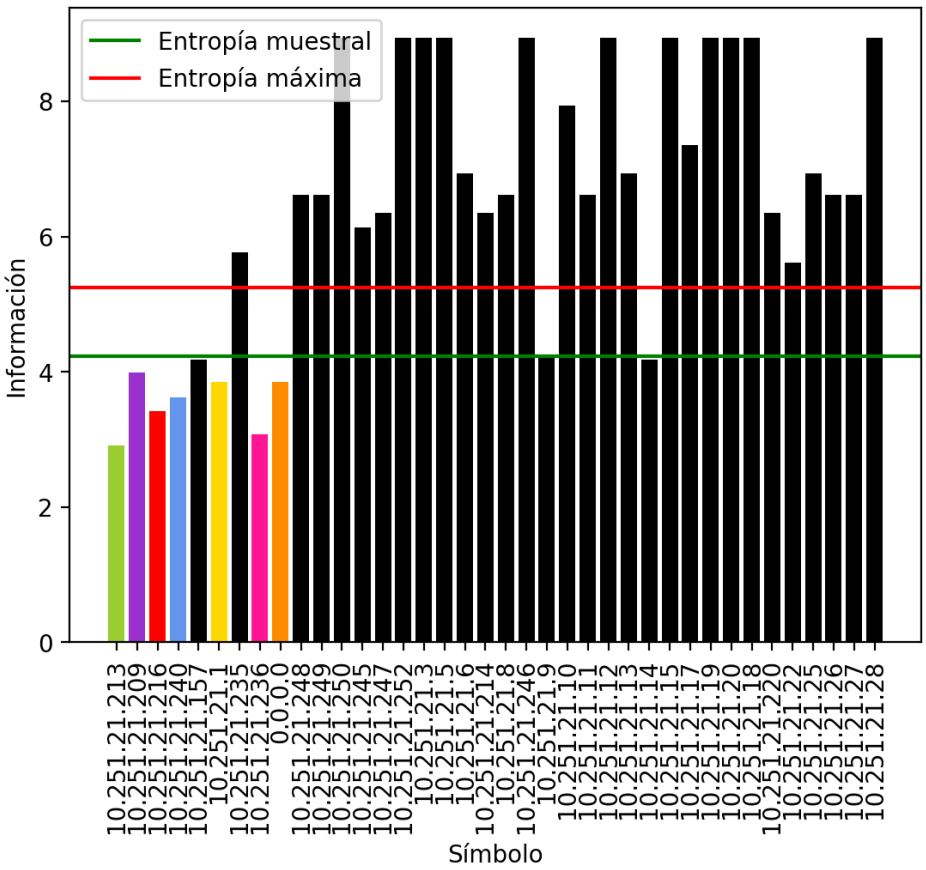

Figura 4: Gráfico de la información de los símbolos de la fuente S_2 observados en esta red. Se muestra la entropía muestral de S_2 y su entropía máxima.

Por último graficamos la red subyacente de mensajes ARP representandola en la figura 5. Cada nodo del grafo representa una IP interviniente en al menos un mensaje ARP de la red, y cada eje orientado representa que un nodo (con su correspondiente Ip) envió al menos un mensaje con Ip destino correspondiente al nodo apuntado, sin distinguir entre paquetes de tipo who-has o is-at.

Dada esta representacion, es importante leer correctamente el grafico, ya que los nodos con mayor grado no son necesariamente los de mayor probabilidad. Dada una Ip origen, puede haber un caso en que mande una gran cantidad de mensajes a unas pocas direcciones destino (recordar que en esta fuente se cuentan la cantidad de mensajes enviados osea las aristas que salen de cada nodo). Como este grafo no posee peso en las aristas, estos intercambios resultarian en un nodo con pocas aristas, mientras que otro intercambio de mensajes donde una ip destino manda comparativamente menos paquetes pero a una mayor cantidad de IPs distintas, resultara en un nodo con muchas mas aristas. Teniendo en cuenta esto podemos notar que igualmente las direcciones con menor informacion (osea mayor probabilidad) resultan las de mayor grado en el grafo.

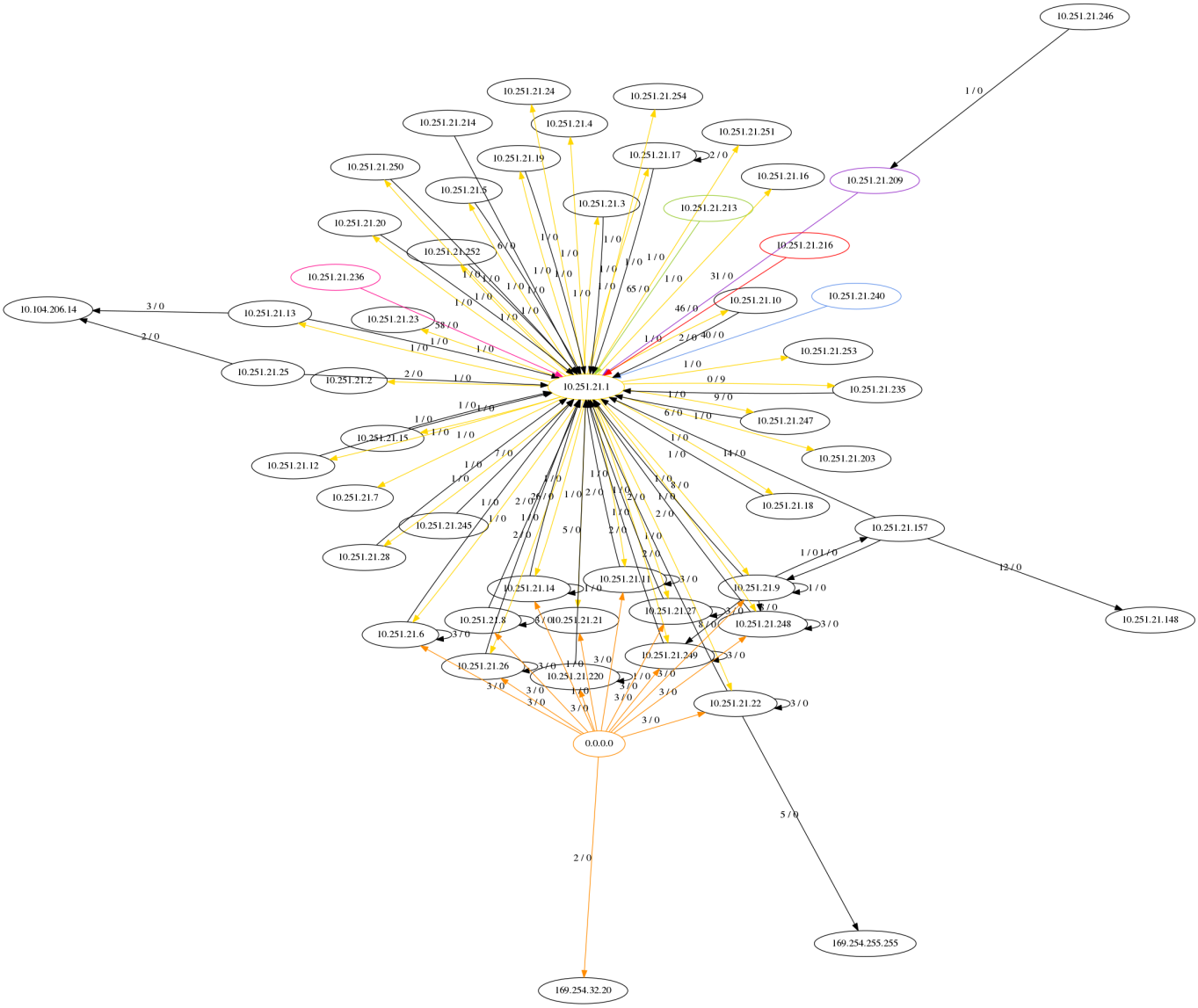

Figura 5: Grafo de la red de mensajes ARP subyacente. Los nodos son las IPs observadas y los ejes son los mensajes ARP. En colores se marcan los nodos distinguidos (información por debajo de la entropía) y sus mensajes salientes. Cada arista tiene anotada la cantidad de requests/replies ARP.

4. Experimento 2: Red hogareña

4.1. Descripción del contexto

El experimento fue realizado en una red doméstica, por medio de una conexión Wi-Fi de Fibertel. Al momento de tomar las mediciones estaban conectados a la red una laptop, una smart TV, un celular y una tablet. La fecha de la captura fue Sábado 7 de Octubre de 2017.

Adicionalmente, el celular estaba tambien siendo usado como control remoto por el programa VLC.

4.2. Descripción de la captura

La captura resulto en un trafico de 10000 paquetes en total. La figura 6 muestra la distribución de protocolos dentro de la captura. Observamos paquetes de tipo IPv4, IPv6(0x86dd), ARP y también Fast Roaming Internet Request. Este último es un protocolo usado en redes inalambricas, especificamente con dispositivos mobiles en movimiento como son los celulares.

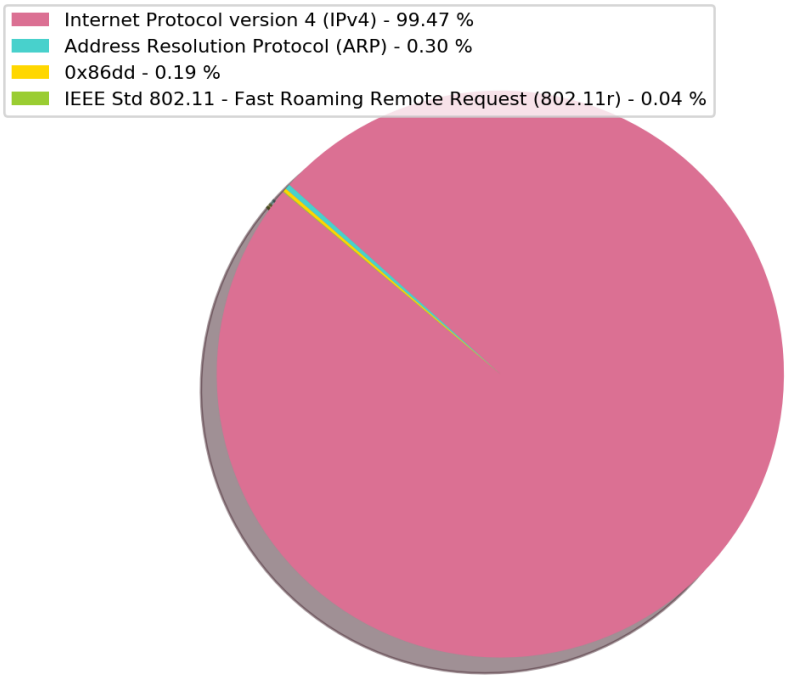

Figura 6: Gráfico que muestra la distribución de protocolos en la red.

En la figura 7 podemos ver el porcentaje de paquetes broadcast sobre el total de paquetes capturados. Vemos que representan un 1,8

En la figura 8 vemos que los protocolos que presentan paquetes de tipo broadcast son ARP e IPv4. El porcentaje de broadcast es mucho menor al de la primer red. Atribuimos esto al hehco de que hay menos dispositivos en uso conectados a la red.

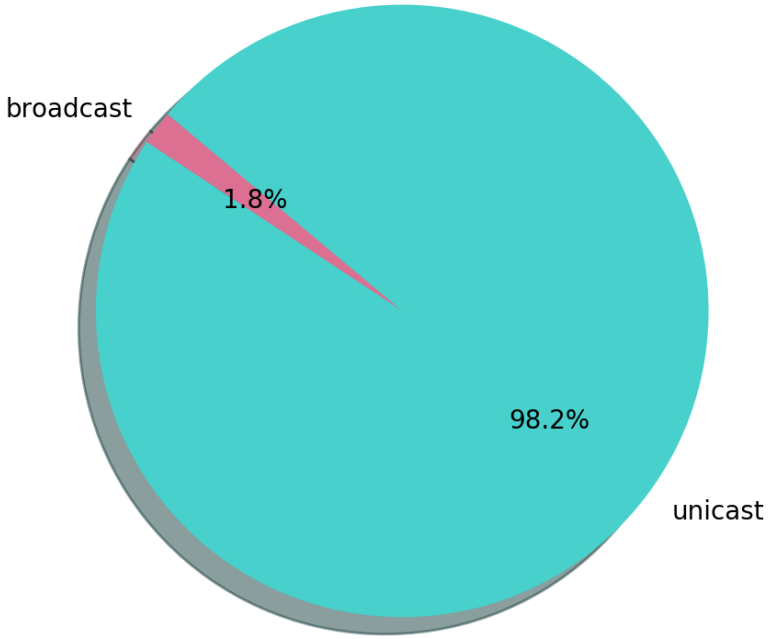

Figura 7: Gráfico que muestra los porcentajes de tráfico broadcast y unicast.

4.3. Análisis de la captura

La figura 8 muestra la información de cada símbolo de la fuente S1 para esta red. Observamos que hay un símbolo cuya información es disitntivaente mas baja en comparación con la de los demás (IPv4 unicast); lo cual hace que la

entropía muestral sea mucho menor a la máxima.

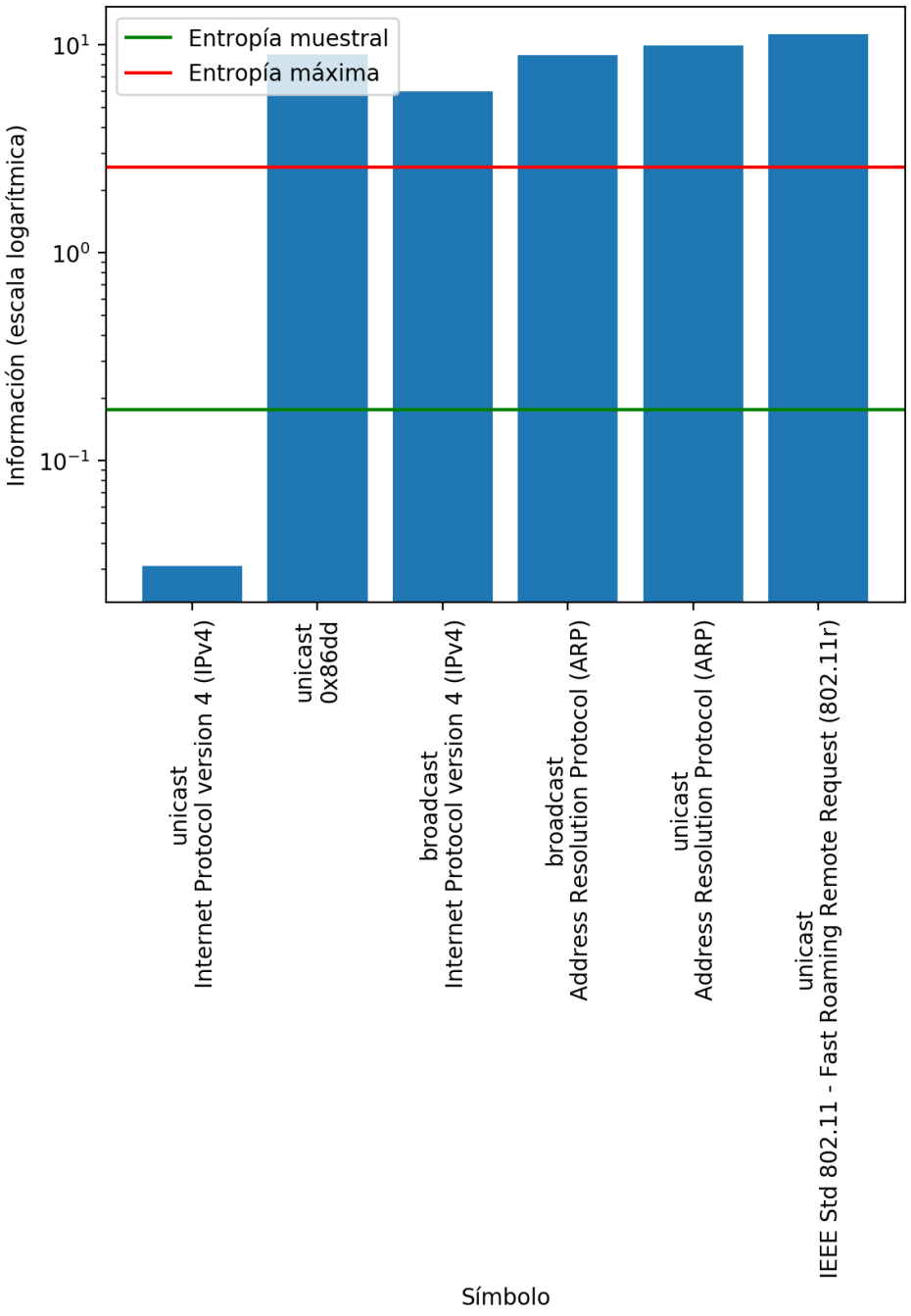

Figura 8: Gráfico de la información de los símbolos de la fuente S_1 observados en esta red. Se muestra la entropía muestral de S_1 y su entropía máxima.

En el caso de la fuente S_2, como podemos ver en la figura 9, hay dos IPs particulares que aportan menos información que el resto. Se trata de las que envian requests de ARP con mas frecuencia. En este caso, al aportar todos los símbolos aproximadamente la misma información (comparativamente con la figura 8 donde hay un simbolo con muy poca informacion) , la entropía muestral se encuentra cerca de la máxima.

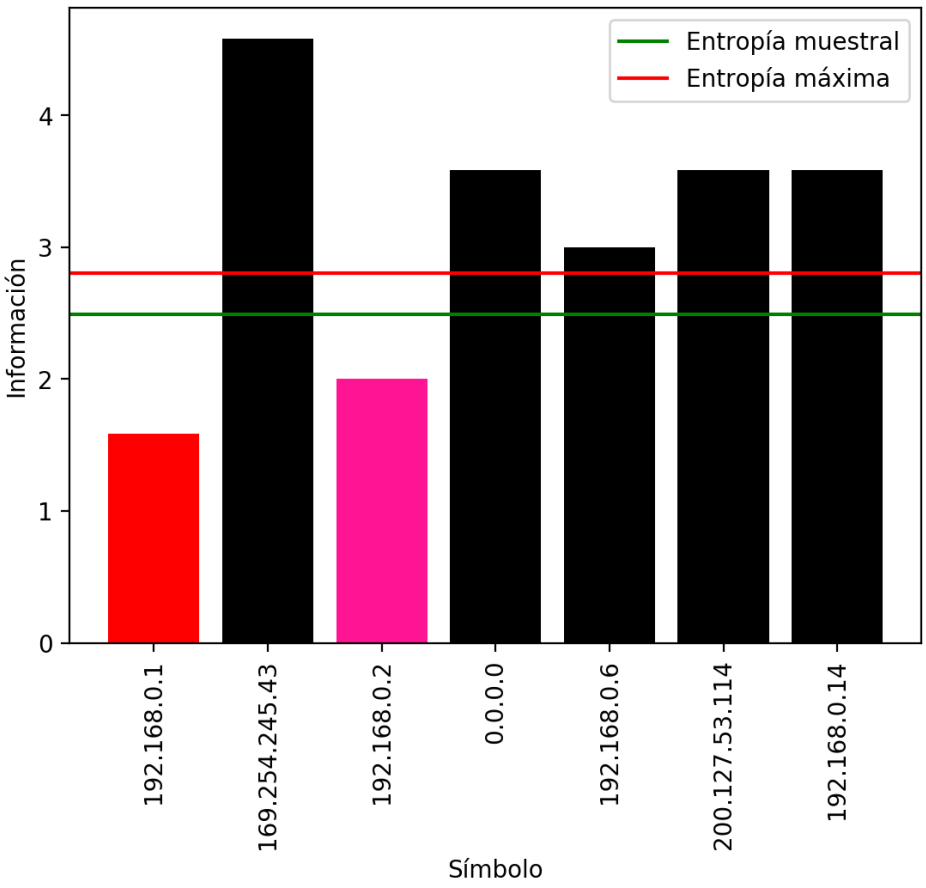

Figura 9: Gráfico de la información de los símbolos de la fuente S_2 observados en esta red. Se muestra la entropía muestral de S_2 y su entropía máxima.

Graficamos la red subyacente de mensajes ARP en la figura 10. Los nodos representan a los hosts y las aristas los mensajes ARP de los dos tipos. Podemos ver que los nodos correspondientes a las dos IPs con menor informacion son las que tienen mayor grado en el grafo (grado 4 para 192.168.0.1 y grado 3 para 192.168.0.2)

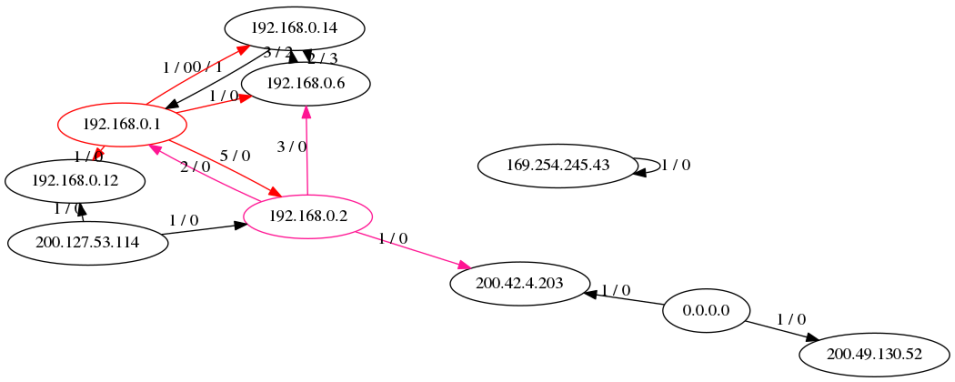

Figura 10: Grafo de la red de mensajes ARP subyacente. Los nodos son las IPs observadas y los ejes son los mensajes ARP. En colores se marcan los nodos distinguidos (información por debajo de la entropía) y sus mensajes salientes. Cada arista tiene anotada la cantidad de requests/replies ARP.

5. Experimento 3: Red hogareña cableada

5.1. Descripción del contexto

El experimento fue realizado en una red doméstica, por medio de una conexión Wi-Fi. Al momento de tomar las mediciones estaban conectados a la red varias laptops y celulares. La fecha de la captura fue Domingo 8 de Octubre de 2017.

5.2. Descripción de la captura

Capturamos 11000 paquetes. En la figura 11 se muestra la distribución de los protocolos en la red. Observamos que la mayoría de los paquetes son de tipo IPv4, y muchos menos son de tipo ARP y `0x86dd` (IPv6).

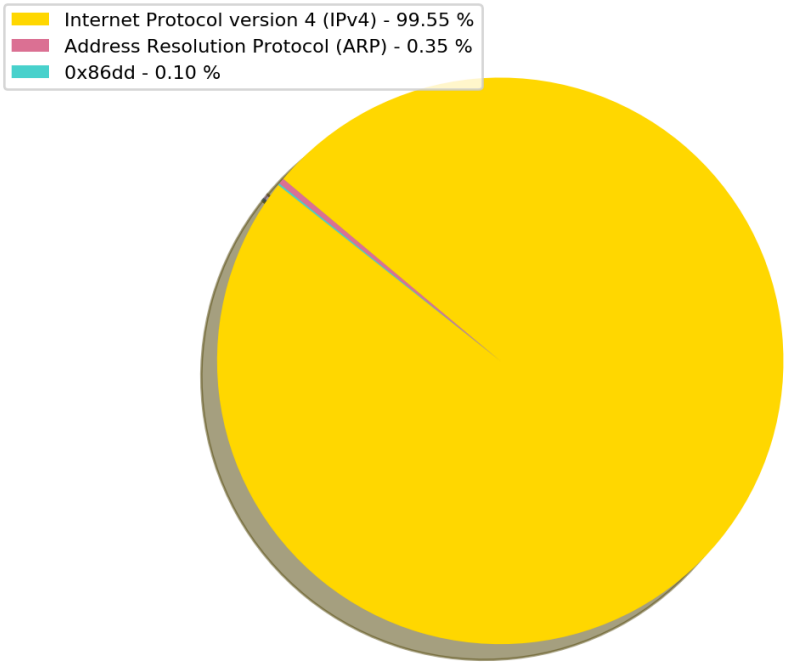

Figura 11: Gráfico que muestra la distribución de protocolos en la red.

En la figura 12 podemos ver el porcentaje de paquetes broadcast comparado con el total de paquetes. Vemos que esto representa un 2,2 % del total. Además, en la figura 13 vemos que los paquetes de broadcast son de tipo IPv4 y ARP.

Este comportamiento es similar al del segundo experimento, ya que ambas son redes domésticas con pocos dispositivos conectados.

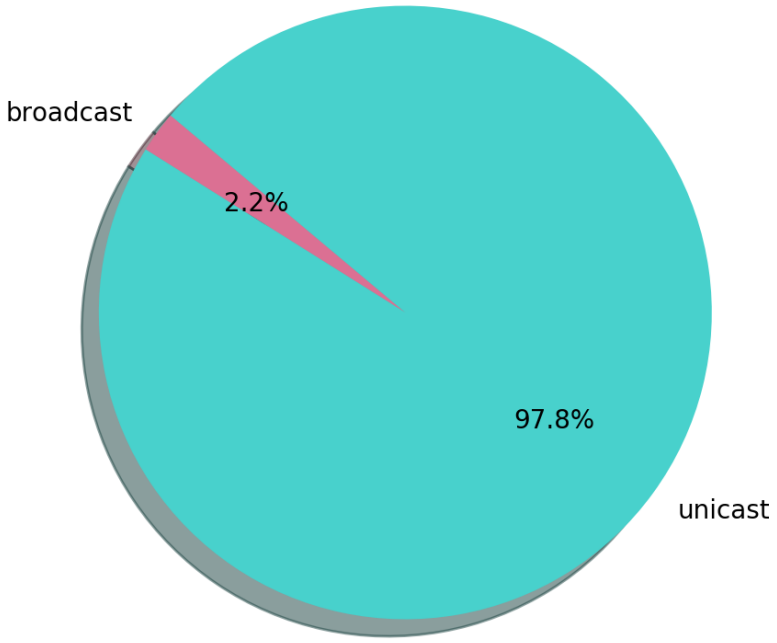

Figura 12: Gráfico que muestra los porcentajes de tráfico broadcast y unicast.

5.3. Análisis de la captura

En la figura 13 se muestra la información de cada símbolo de la fuente S1. Hay un símbolo con mucha menor información que los demás (IPv4 unicast), lo cual se debe a que la mayoría de los paquetes fueron de este tipo. A causa de esto, la entropía de la fuente es muy baja y está muy lejos del máximo, debido a que hay grandes diferencias entre el símbolo de menor información y los demás.

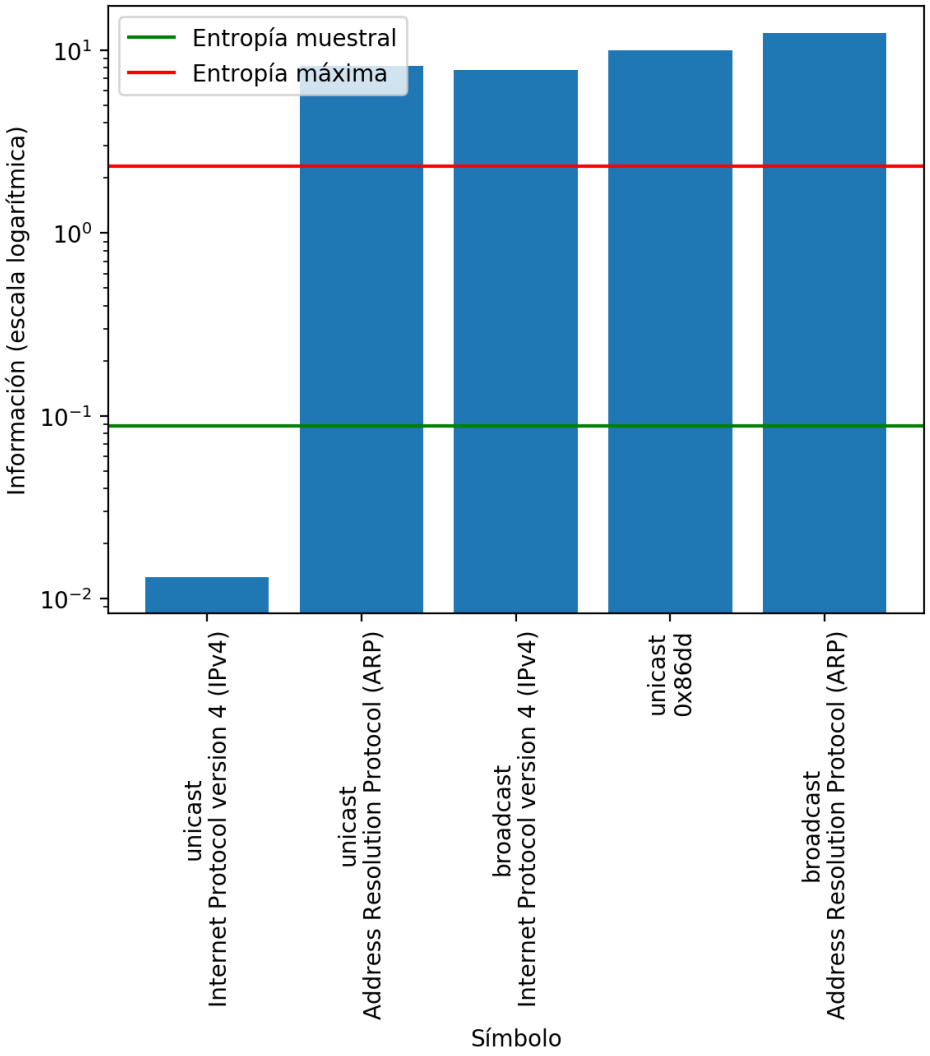

Figura 13: Gráfico de la información de los símbolos de la fuente S_1 observados en esta red. Se muestra la entropía muestral de S_1 y su entropía máxima.

En cuanto a la fuente S_2, vemos en la figura 14 que sólo hay dos símbolos correspondientes a las dos direcciones IP que realizan requests ARP. Atribuimos esto a lo reducida que es la red. La entropía muestral está cerca de $1/2$, la mitad que la entropía máxima.

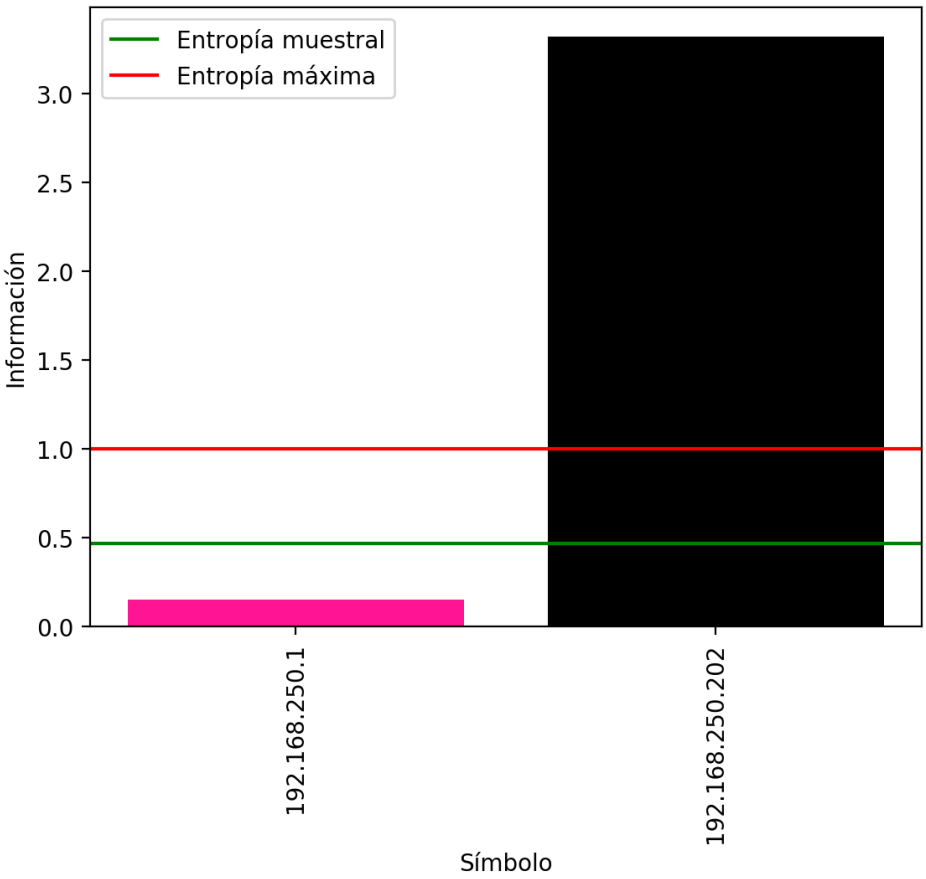

Figura 14: Gráfico de la información de los símbolos de la fuente S_2 observados en esta red. Se muestra la entropía muestral de S_2 y su entropía máxima.

Por último representamos los envios ARP dentro de la red durante el periodo monitoreado, al igual que en los otros dos experimentos. Notese que aqui se representan tanto los requests ARP como sus respuestas, por lo que aparecen aqui cuatro nodos a diferencia de las dos direcciones que aparecen com simbolos de la fuente. Al ser una muestra sobre una red tan pequeña este grafo no muestra muchos mas aspectos relevantes para el analisis.

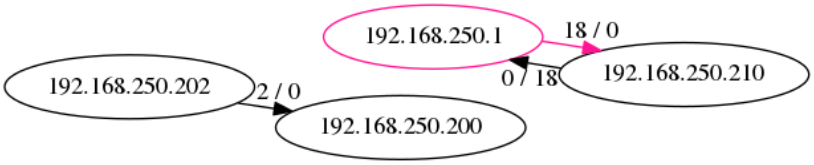

Figura 15: Grafo de la red de mensajes ARP subyacente. Los nodos son las IPs observadas y los ejes son los mensajes ARP. En colores se marcan los nodos distinguidos (información por debajo de la entropía) y sus mensajes salientes. Cada arista tiene anotada la cantidad de requests/replies ARP.

6. Conclusiones

Lo primero que hemos notado es lo cuidadosos que debemos ser al diseñar una fuente de Informacion y sobre todo la manera de agrupar y/o distinguir los simbolos, ya que estas decisiones pueden resultar en sets de datos muy diferentes.La fuente S1 sirvio para agrupar por protocoles, mientras que la S2 para distinguir direcciones de quienes mandaban requests ARP. En cuanto a como cambia los resultados una desicion al diseñar la fuente, tenemos por ejemplo que en el tercer experimento dos direcciones que mandaron respuestas a las requests ARP y por tanto no aparecian como simbolos de S2. Si hubieramos diseñado S2 de tal manera que las respuestas a requests ARP tambien son consideradas simbolos, contariamos con sets de datos distintos y la entropia se hubiese visto modificada.

Observamos tambien que la entropia funciona a nivel practico como una cota bajo la cual se encuentran los simbolos mas relevantes como se ve claramente en el segundo experimento donde lo central del analisis se refiere exclusivamente a los simbolos que se encuentran por debajo de la entropia.

Otra cosa interesante de ver fue la brecha entre la entropia muestral y la maxima, ya que nos resulto una manera sencilla de cuantificar lo bueno o malo del diseño de una fuente, o lo atipico de una medicion en funcion a cuanto se acercaba a la entropia maxima. Dicho en otras palabras, la entropia maxima representa la entropia para una fuente equiprobable (donde la probabilidad de cada simbolo es uniforme).

De los tres experimentos el mas interesante parecio ser el primero, ya que se trataba de una red abierta inalambrica con gran cantidad de dispositivos conectados a la misma. Por eso presento mayor variabilidad de datos y un analisis mas rico.

En general concluimos que la teoria de la informacion es una herramienta util para el analisis ya que permite ordenar claramente los datos y agruparlos adecuadamente para el analisis que se quiera realizar, ademas de proveer calculos y resultados teoricos como la entropia para poder sacar conclusiones.