DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

TP1: Wiretapping

Primer cuatrimestre de 2019 — Teoría de las Comunicaciones

Integrante	LU	Correo electrónico
Goldstein, Brian	027/14	brai.goldstein@gmail.com
Gamarra, Ignacio	792/11	gamarra_mi32@yahoo.com.ar
Delgadillo, Abel	74/12	adelgadillo91@gmail

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) 4576-3359
http://www.fcen.uba.ar

Índice

1. Introducción

En el presente documento se exponen algunas técnicas y herramientas provistas por la Teoría de la Información para estudiar las diversos aspectos del tráfico de las redes, capturando y analizando los paquetes de datos que circulan en las mismas.
Para esto desarrollaremos aplicaciones que nos permitan capturar dichos datos y analizar estos paquetes utilizando herramientas de manipulación y análisis de paquetes recomendados por la cátedra (Wireshark y Scapy).

1.1. Marco Teórico

Antes de introducir los experimentos realizados explicaremos los conceptos involucrados.
En los experimentos analizaremos aspectos de la **capa de enlace** (segunda capa del Modelo OSI), la cual es responsable de la transferencia de información a través de un circuito de transmisión de datos. Recibe las peticiones de la capa superior (capa de red) y utiliza los servicios de la capa física. Esta capa es la facilidad de área extensa por la que se pueden comunicar los sistemas mediante un protocolo de la capa de enlace de datos.

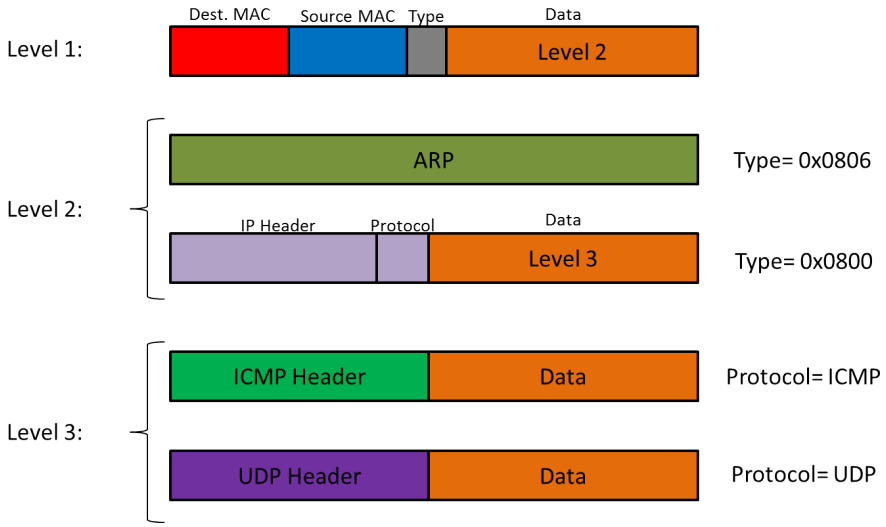

Figura 1: Interfaces de Comunicación Ethernet

Entre estos protocolos tenemos ARP (Address Resolution Protocol) que se encarga de ser el nexo entre la capa de enlace y la capa de red. Su principal responsabilidad es asociar direcciones IP con una dirección MAC (Media Access Control), más precisamente, encuentra la dirección MAC que corresponde a una IP dada.
Para ello, cada vez que un Host necesita comunicarse con otro y su dirección primero emite un mensaje de Broadcast en la red, consultando quién posee una IP específica y después espera la respuesta de alguna de los Host conectados a la red.
Los paquetes ARP pueden ser un request de tipo **Broadcast** (*Who has*) o un reply **Unicast** (*Is at*). En el request de tipo Broadcast se coloca la MAC de destino con el valor `FF:FF:FF:FF:FF:FF` y se pregunta cuál host tiene una dirección IP específica. En la respuesta Unicast la dirección MAC destino es del host quien consulto por la IP específica y la dirección MAC de origen tiene el valor de la IP preguntada.

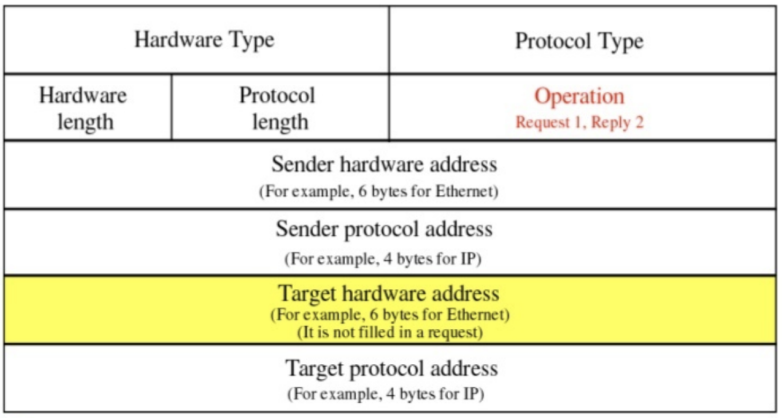

Figura 2: Formato ARP

De esta manera cada Host completa una **tabla** con las distintas direcciones MAC e IP relacionadas y así poder enviar paquetes en la red.

Con esto en mente, modelaremos la fuente de estos mensajes basándonos en la Teoría de la Información de Shannon, midiendo la probabilidad de cada símbolo, cantidad de información y su entropía. A continuación, una breve introducción a estos conceptos teóricos:

Una **fuente de memoria nula** emite símbolos los cuales aportan una cantidad de información determinados por el siguiente cálculo:

$$I(s) = -log(P(s)) \tag{1}$$

Donde $P(s)$ es la probabilidad de aparición del símbolo s en la fuente S.

Este valor es necesario para luego calcular la **Entropía** de la fuente con la siguiente fórmula:

$$H(S) = \sum_{s \in S} P(s)I(s) \tag{2}$$

Esto nos proporciona una métrica de la cantidad de información esperada que emite la fuente, y luego podremos detectar los símbolos distinguidos que disten más de la entropía. Los que aporten más valor en consecuencia aportan más información de lo esperado y los de menor valor a su vez representan menos a la información, debido a su alta frecuencia.

Además, nos resulta interesante analizar la probabilidad de los símbolos para ver cuán distante se encuentra la entropía de las fuentes a proponer de la entropía máxima teórica, la cual se da cuando los símbolos son equiprobables, dando por resultado:

$$H(S)_{max} = I(s) \tag{3}$$

Con estas bases teóricas podemos establecer un criterio para considerar que un nodo de la red es un **nodo distinguido** para una fuente: un nodo es distinguido cuando la información asociada a la aparición de su IP es menor que la entropía de dicha fuente.

2. Métodos y Condiciones de los Experimentos

Hemos realizado 3 experimentos en 3 diferentes redes respectivamente. 2 de ellas corresponden a redes inalámbricas Wifi y una a través de una conexión cableada en un switch.

2.1. Capturas

Para la captura de los paquetes se ha programado en código Python una herramienta que utiliza la librería **Scapy** [1] de Python. También usamos la herramienta **Wireshark** [2] para la captura de paquetes, almacenando las capturas en formato `pcap` para luego ser analizadas. Usando la función `sniff()` provista por Scapy pudimos capturar tráfico de distintas redes locales y así representar las fuentes modeladas:

- $S_1 = \{s_1, s_2, ..., s_q\}$, la cual es la fuente de información donde cada elemento s_i es una tupla formada por el tipo de destino de la trama (unicast o broadcast) y el protocolo asociado (ARP, IP, IP6, IPX, PPP, CDP, PPPoE, DTP), establecido en el enunciado.

  Para esto utilizamos la función `sniff()` estableciendo la cantidad total de paquetes a capturar. Luego para cada uno vemos según el destino del paquete si se trata de un request tipo Broadcast o Unicast (chequeando si el destino se corresponde con el valor `ff:ff:ff:ff:ff:ff` establecido por el protocolo). Luego contamos las apariciones de cada uno de estos símbolos para posteriormente calcular probabilidad, entropía e información.

- $S_2 = \{s_1 \dots s_n\}$, determinada por nosotros, donde $s_i =< IP\_src_i, cast\_type_i >$ son las direcciones IP de origen de los paquetes ARP *who-has*. Esta fuente nos permite identificar las interfaces de red emitiendo información sobre el medio, para posteriormente detectar los nodos distinguidos de la red en cada captura. Como habíamos mencionado, un nodo es distinguido cuando la información asociada a la aparición de su IP es menor que la entropía de la fuente.

  Para capturar paquetes con este nuevo modelo implementamos un código usando la función `sniff` mencionada, pasando como parámetro un filtro para obtener sólo los paquetes que usan el protocolo ARP. De los paquetes recibidos contamos la cantidad de veces que cada direccion IP aparecía como origen y como destino.

  Finalmente para esta fuente, generamos los datos para construir un grafo representativo de la red capturada tomando como nodos las direcciones IP que figuran en ella. De esta forma podemos conocer la topología de la red analizada.

Con estos datos calculamos en nuestra implementación la probabilidad y cantidad de información para cada símbolo, y finalmente la entropía de la fuente y su entropía máxima teórica.

Realizamos 3 capturas de tráfico con este código donde cada una de ellas está formada por 10000 tramas para identificar en cada una de las capturas los símbolos que más información proporcionen y así encontrar los nodos distinguidos.

2.1.1. Wework

Esta medición fue realizada en el edificio corporativo Wework, un día Lunes en el horario de las 12:00pm durante 2 horas aproximadamente utilizando el código mencionado anteriormente.

[1] `http://www.secdev.org/projects/scapy`
[2] `https://www.wireshark.org/`

La conexión fue mediante Wifi en la cual contemplan la conexión de computadoras, 2 impresoras y varios teléfonos celulares. No fue posible calcular la cantidad exacta de equipos conectados pero en la misma red estimamos que habían alrededor de 200 computadoras y 3 access points ya que la red se compartía entre 3 pisos del edificio.

2.1.2. Casa

Estas capturas fueron realizadas en una red LAN hogareña, un día Lunes en el horario de las 21:00 hs, por un lapso de 1 hora. La herramienta utilizada fue la mencionada anteriormente. En esta LAN habían conectadas 3 computadoras y 3 celulares al momento de la medición mediante conexión Wifi.

2.1.3. Escuela

La conexión en este caso fue cableada mediante un switch de una red en un laboratorio de una escuela, un día martes alrededor de las 14hs. El switch proporcionaba conexión a otros equipos en el piso. Si bien habían alrededor de 30 equipos conectados, algunos de ellos no estaban encendidos al momento de realizar la medición, con lo que obtuvimos información de aproximadamente 20 equipos conectados.

3. Resultados

3.1. Wework

Daremos ahora análisis a la capturas de la red wework. Esta resulto ser la red más grande con la que trabajamos, comenzaremos analizando el tráfico por protocolo y la distinción entre Broadcast y Unicast, luego procederemos al análisis de las fuentes S1 y S2 y sus respectivas propiedades descriptas por la teoria de la información.

3.1.1. Tráfico por protocolo:

A continuación podemos ver los resultados de la medición de tráfico por tipo de protocolo:

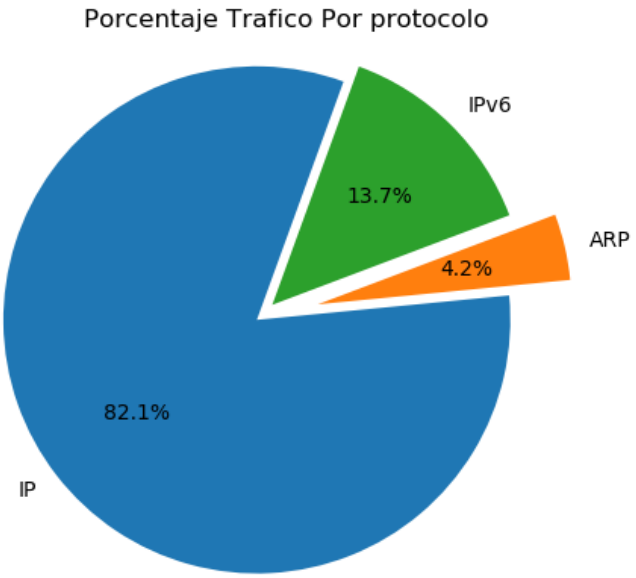

Figura 3: Tráfico por protocolo al de las capturas con las que se genera S1

Se observan 3 tipos de protocolos distintos `IP`,`IPv6` y `arp` todos protocolos esperados y estudiados en la materia.
Se observa que la mayor parte de los paquetes son de los protocolos `IP`, `IPv6`. Esto se debe a que hay mayor movimiento de datos de usuario que de control. Se observa que si bien la gran mayoría de paquetes (poco mas de 95 %) son de `IP` o `IPv6` la carga de los paquetes `ARP` es no despreciable, tomando un 4 % de la red solo para resolver la comunicación de capa 2.
Se observan las proporciones en las que circulan paquetes de `IPv6` en relación a `IP`, siendo loas primeros 6 veces (aproximadamente) la cantidad de los segundos. Es un aspecto a destacar que si bien hay mayor probabilidad del símbolo IPv4, se está comenzando a implementar el uso de protocolo IPv6 a gran escala.

3.1.2. Tráfico Unicast-Broadcast:

Se muestra a continuación los porcentajes que indican el volúmen de paquetes de cada categoría (unicast o broadcast) independientemente de los protocolos usados. Para esto basta con ver el destino y analizar si es o no dirección de broadcast:

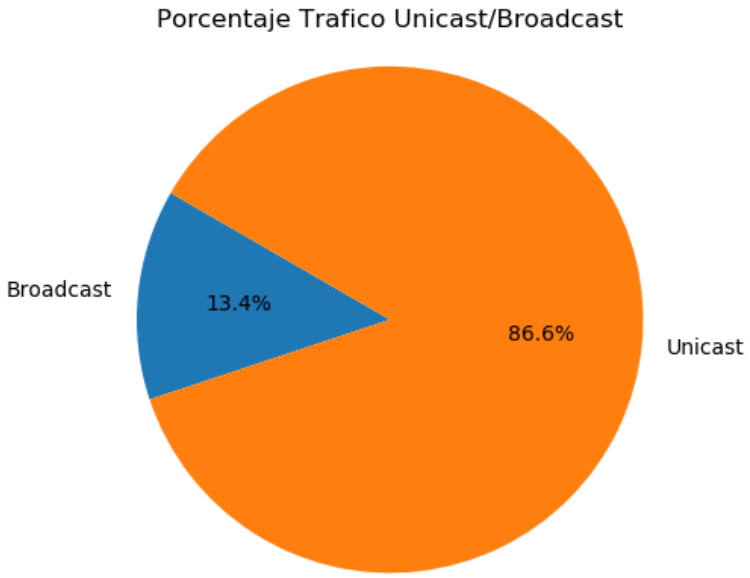

Figura 4: S1: Wework - Tráfico según tipo de destino de trama

Podemos notar según la Figura 4 que la cantidad de paquetes de **Broadcast** es muy significativa, esto seguramente significa un sobrecargo importante en la eficiencia de la red por dos motivos, el primero es que los paquetes de Broadcast congestionan mas la red ya que deben ser enviados a todos los nodos de la misma, la segunda es que estos paquetes suelen ser de control por lo que serían netamente overhead.

3.1.3. Analisis de S1:

Analizaremos las propiedades de la fuente S1 pedida con las herramientas provistas por la teoria de la información. El gráfico a continuación mostrará la cantidad de información por símbolo comparándola con la entropía de la fuente y la entropía máxima teórica, hacemos especial distinción entre símbolos provenientes de distintos protocolos y símbolos con el mismo protocolo pero diferente categoría de destino(Broadcast-Unicast).

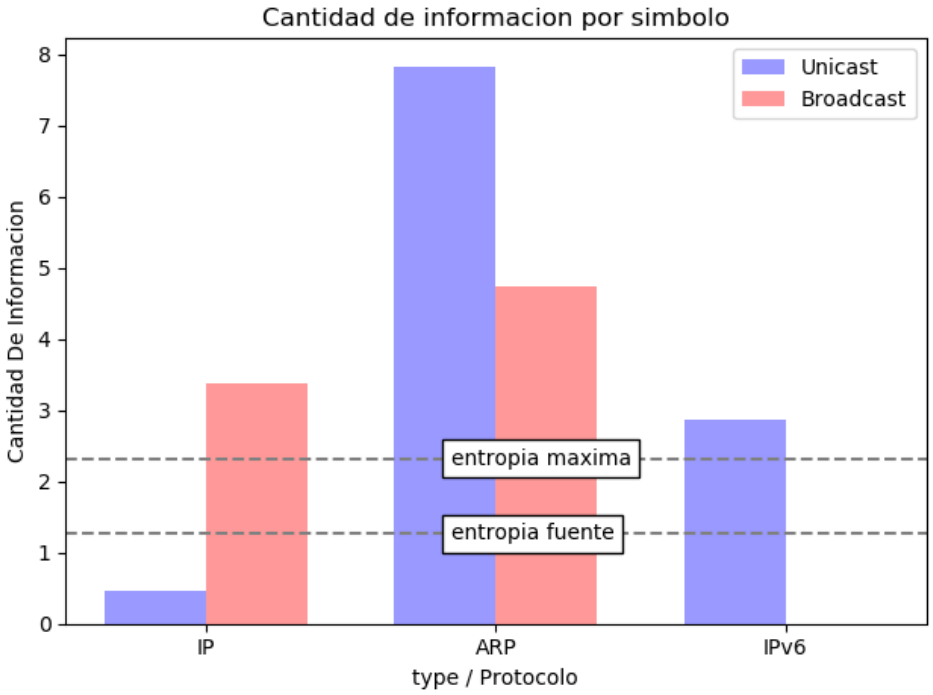

Figura 5: S: Wework -Cant. información según símbolo distinguidos por Broadcast/Unicast

Se observa del gráfico como de los 3 protocolos los símbolos que aportan mas información son aquellos que son de Broadcast o de ARP. Le atribuimos este comportamiento a la gran desproporción de paquetes IP (unicast) que circulan, esto tiene el efecto de subir mucho la probabilidad del símbolo $< IP, Unicast >$ y por consecuente bajar la probabilidad de todos los otros símbolos. De no tenerse en cuenta los paquete IP de unicast , se observa que los símbolos $< IP, Broadcast >$ y $< IPv6, Unicast >$ están cerca de la entropía máxima teórica, esto indica que sus información esta cerca del nivel de distribución óptima de la información.
Se observa la entropía real de la fuente sustancialmente por debajo de todos los símbolos menos $< IP, Unicast >$ esto es esperable dado que al haber tantos de estos últimos (y al ser la entropía definida como la esperanza de la información) es razonable que este símbolo de muy baja información baje el promedio (en el sentido de esperanza) de la fuente.
Por último observamos que ambos símbolos de ARP son distinguidos por ser de muy alta información, esto es debido a la poca frecuencia relativa de estos.

3.1.4. Análisis de S2:

Como fue explicado en la sección de metodologías nuestra fuente S2 planteada consta de las IPs de los diferentes nodos de la red. Estas fueron capturadas desde los paquetes ARP (como fue explicado en esa sección). En esta sección analizaremos las propiedades desde el punto de la teoria de la información de esta fuente generada.
Para el análisis de la información de fuente S2 graficaremos la información de los diferentes símbolos:

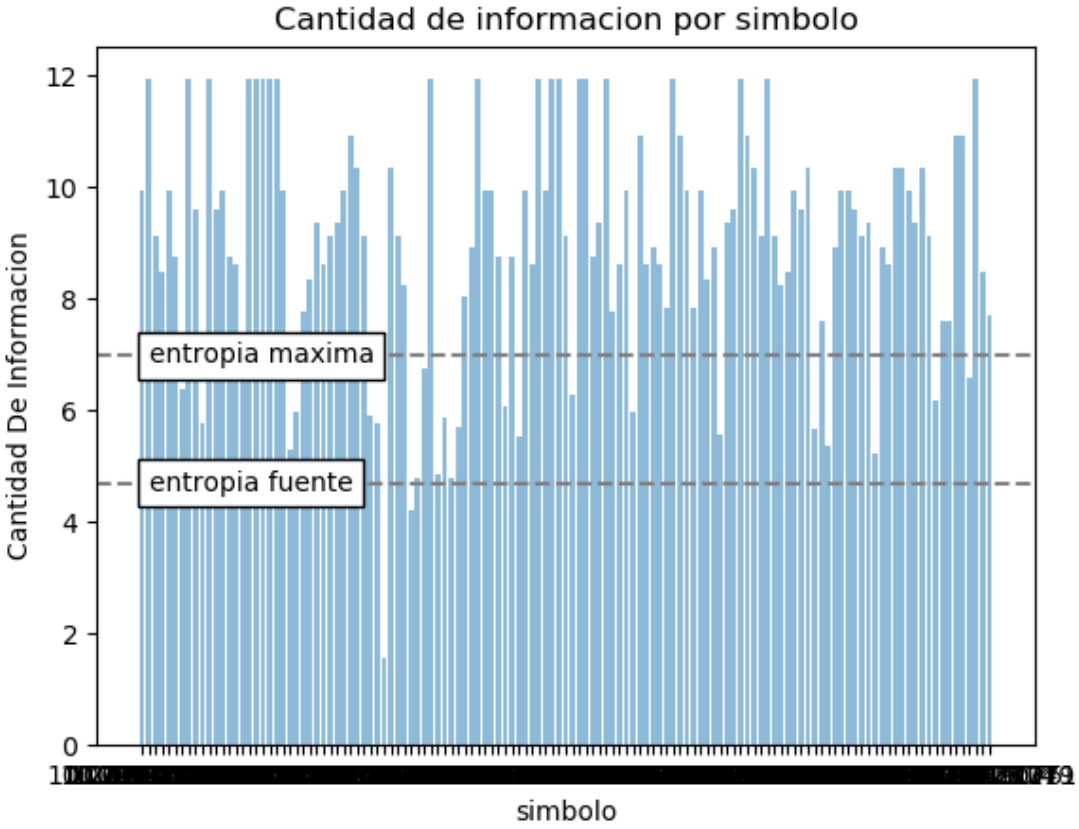

Figura 6: Wework-**Entropía de la fuente:** 1.38175646394 ; **Entropía máxima:** 6.98868468677

Se observa cómo la gran mayoría de símbolos están por arriba de la línea de entropía (real) de la fuente, esto nos indica que de nuevo, hay unos pocos símbolos que son repetidos con mucha frecuencia. Esto se explica con entender que la mayoría de computadoras de la red usualmente necesitan comunicarse directamente con el router (o algun access point) y estas IP son las más repetidas en las capturas. A tal punto que solo este símbolo desbalancea la entropía de la fuente disminuyéndola considerablemente (ya que la capacidad de predecir un símbolo es alta, altamente probable el símbolo sea la IP de router).
Se observa también que hay una gran cantidad de símbolos con una cantidad de informacion que supera considerablemente la entropía máxima, esto parece indicar (en contraste con el gran

número de símbolos por debajo de la entropía máxima) la disparidad del uso de la red de las diferentes computadoras. Desde un punto de vista teórico estos serían símbolos distinguidos, empíricamente esto expresa que estas computadoras usan la red de forma mas moderada.

Podemos notar según la Figura 4 que la cantidad de paquetes de Broadcast es muy significativa.

3.1.5. Topología

Analizaremos la topología de la red, mediante un grafo de conexiones basado en los paquetes ARP. En el resto de las capturas el grafo presentado distingue las diferentes IP y también el sentido en el que fluye la información (grafo dirigido), sin embargo en el caso de la captura de Wework al ser tan extensa la red, no puede visualizarse bien la red con tanta precisión (al menos no de manera estática como es preciso hacer en un informe) nos concentramos en cambio entonces en tratar de estudiar su topología a gran escala y para esto decidimos en este caso no distinguir los nodos ni tampoco el flujo de la información, el resultado es un grafo no dirigido de las comunicaciones capturadas:

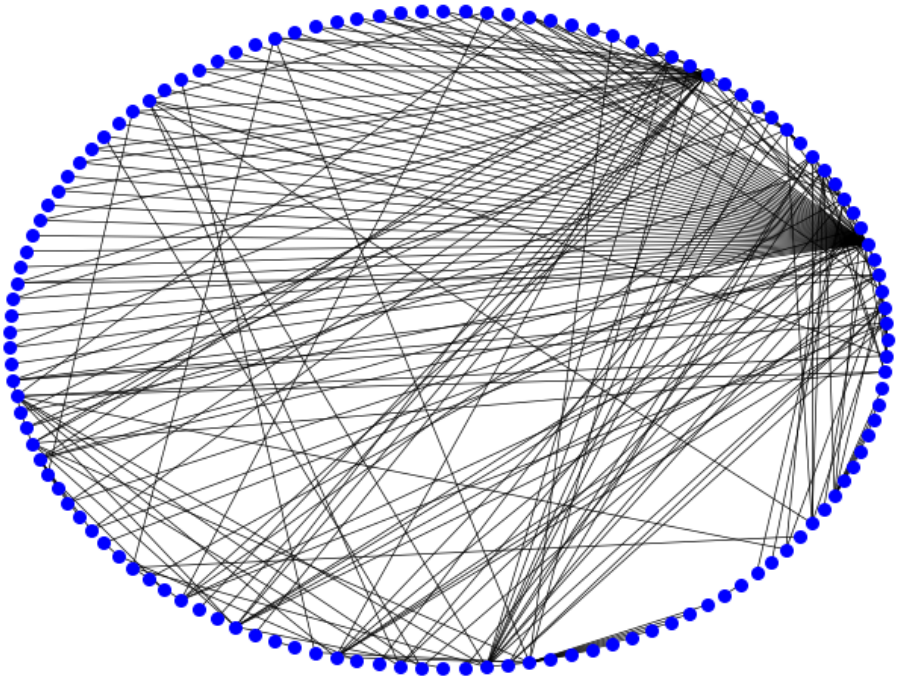

Figura 7: Grafo de red Wework

Se observa en el gráfico la gran cantidad de nodos (son 127). Se puede apreciar como hay nodos con sustancialmente más conexiones, éstos son los que desbalancean la entropía en el análisis anterior, probablemente router y similares puntos de centralización de grandes tráficos. Se observa tambien nodos marginales con relativamente pocas conexiones, estos son los que en el análisis anterior se los denominó como moderados a la hora de usar la red.

3.2. Casa

Daremos ahora análisis a la capturas de la red de Casa. Esta será la red mas pequeña con la que trabajamos, comenzaremos analizando el tráfico por protocolo y la distinción entre Broadcast y unicast, luego procederemos al análisis de las fuentes S1 y S2 y sus respectivas propiedades descriptas por la teoría de la información y terminaremos analizando su topología a partir del análisis de S2.

3.2.1. Trafico por protocolo:

A continuación podemos ver los resultados de la medición de tráfico por tipo de protocolo:

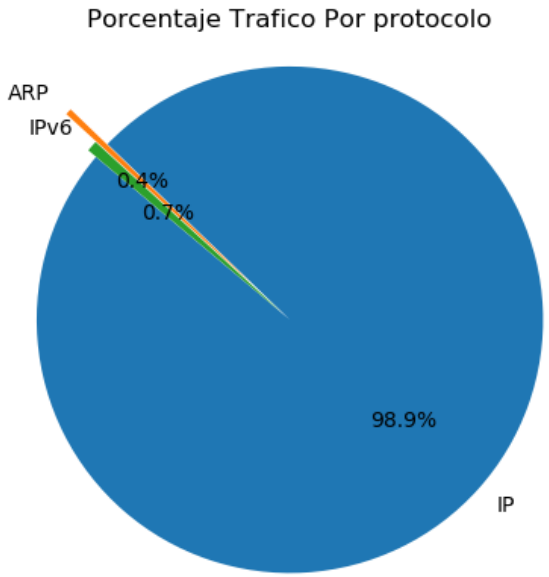

Figura 8: Paquetes por protocolo en una casa

Se observan 3 tipos de protocolos distintos `IP`,`IPv6` y `ARP` todos protocolos esperados y estudiados en la materia.
Se observa que la mayor parte de los paquetes son de los protocolos `IP`. Esto se debe a que hay mayor movimiento de datos de usuario que de control. Se observa que el porcentaje de paquetes ARP es casi despreciable por lo que en esta caso estos no presentan un overhead significativo.
Se observa la poca frecuencia de paquetes `IPv6`, se interpreta de esto que el protocolo IPv4 es suficiente para este tipo de redes y que por tanto dificilmente se vean cantidades significativas de paquetes de este protocolo en redes chicas.

3.2.2. Tráfico Unicast-Broadcast:

Se muestra a continuación los porcentajes que indican el volúmen de paquetes de cada categoría (unicast o broadcast) independientemente de los protocolos usados. Para esto basta con ver el destino y analizar si es o no dirección de broadcast:

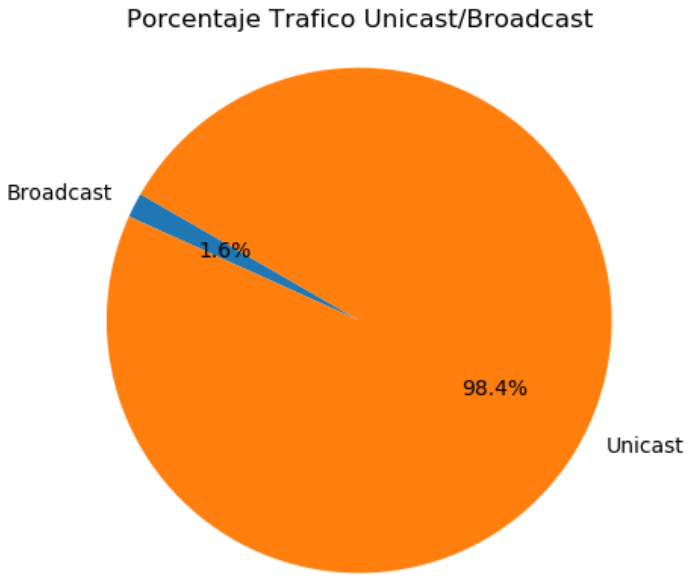

Figura 9: S1: Casa - Torta según tipo de destino de trama

Claramente podemos encontrar una predominancia de paquetes Unicast que se corresponden con los paquetes `IP` vistos en la figura 8. Los de Broadcast son usualmente usados para control de la comunicación por lo tanto son usualmente overhead, es una buena señal de eficiencia que estos representen menos del 2 % . Esto en conjunto con lo observado en el punto anterior parece indicar que el overhead de un red pequena es bajo, La inmensa mayoría de paquetes que viajan por la red son paquetes que cargan información de usuario (o de capas superiores).

3.2.3. Análisis de S1:

Analizaremos las propiedades de la fuente S1 pedida con las herramientas provistas por la teoría de la informacion. El gráfico a continuación mostrará la cantidad de información por símbolo comparándola con la entropía de la fuente y la entropía máxima teórica, hacemos especial distinción entre símbolos provenientes de distintos protocolos y símbolos con el mismo protocolo pero diferente categoría de destino(Broadcast-Unicast).

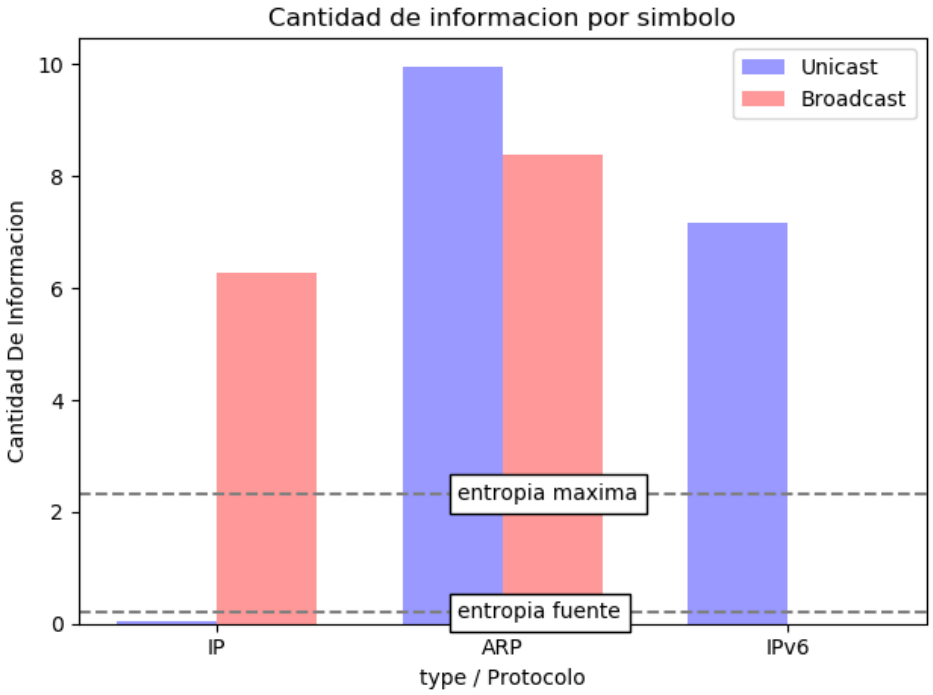

Figura 10: S1: Casa - Cant. información según símbolo

En la figura 10 se puede ver más claramente que la Entropía de la fuente es muy baja, tomando el valor de 0,199535565593. Nos estaría indicando que la incertidumbre de la información es muy reducida, o mejor dicho, los símbolos emitidos son muy previsibles. Esto se da por la aparición predominante del símbolo `IP` que al poseer mayor probabilidad provoca que sea destacado como nodo distinguido.

Se obsverva que ninguno de los símbolos esta cerca a la entropía máxima indicando que la distribución de los símbolos esta muy lejos de ser uniforme. Si no se tiene en cuenta el símbolo $< IP, Unicast >$ la cantidad de información en de los demás símbolos es comparable por lo que sacando este símbolo que desbalancea las entropías la distribución es bastante equilibrada.

3.2.4. Análisis de S2:

Como fue explicado en la sección de metodologías nuestra fuente S2 planteada consta de las IPs de los diferentes nodos de la red. Estas fueron capturadas desde los paquetes ARP (como fue explicado en esa sección). En esta sección analizaremos las propiedades desde el punto de la teoría de la información de esta fuente generada.
Para el análisis de la información de fuente S2 graficaremos la información de los diferentes símbolos:

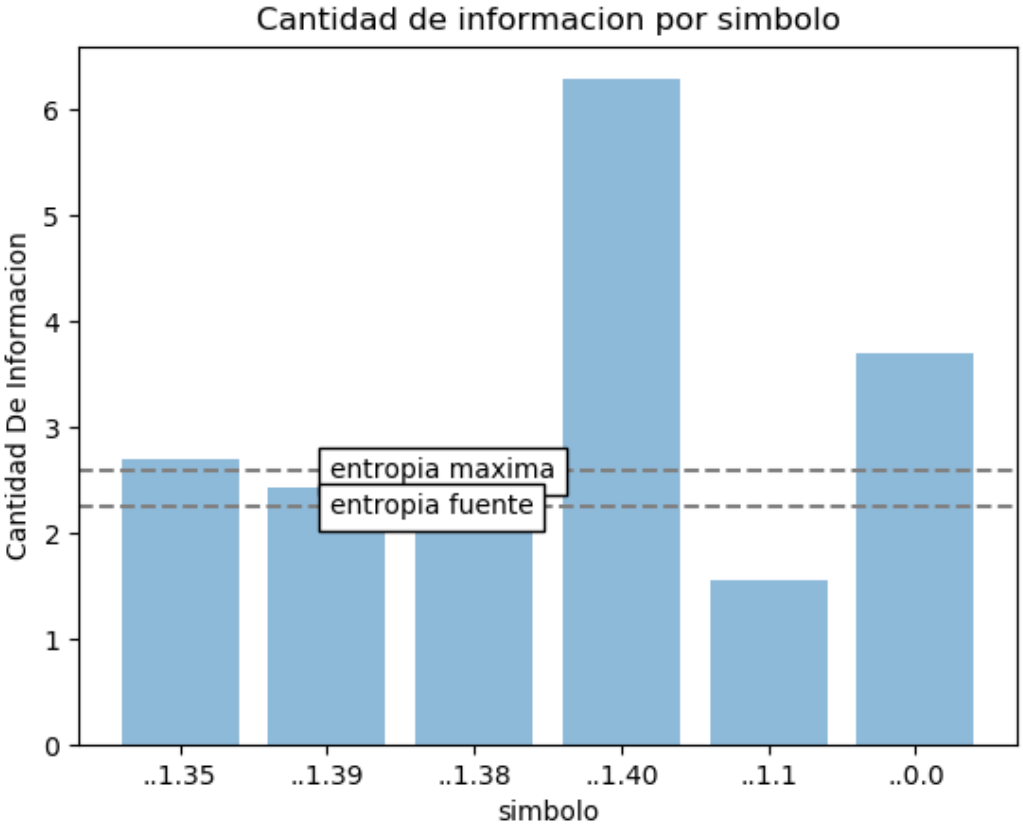

Figura 11: Casa -**Entropía de la fuente:** 0.199535565593 ; **Entropía máxima:** 2.32192809489

Se observa como la entropía de la fuente está cerca a la entropía máxima, indicando así que la esperanza de la información por símbolo sea similar a una fuente ideal con todos los símbolos distribuidos equiprobables.
Se observa tambien que todos los símbolos están cerca de la entropía a excepción 1.40 el símbolo que con menos frecuencia aparece. El símbolo .1.1 es el que con mas frecuencia aparece y el único por debajo de la entropía , esto se explicaria si es el simbolo del router ya que todos los dispositivos se comunican frecuentemente con el por lo que es esperable que este por debajo de la entropía.

3.2.5. Topología

Analizaremos la topología de la red, mediante un grafo de conexiones basado en los paquetes ARP. Planteamos un grafo dirigido donde se captura en los paquetes ARP las direcciones de origen y destino(el pedido en who-has) en los who-has y dibujamos un nodos por cada símbolo de S2 (cada IP) y un eje de origen a destino.

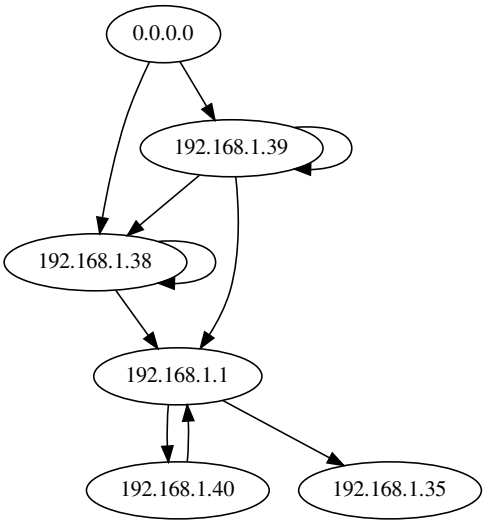

Figura 12: Grafo de red Casa

Podemos notar como fue observado en el punto anterior que el símbolo 192.168.1.1 es el que más flujo de datos controla. Se observa como la red es mucho más chica que la red de wework analizada antes lo que explica que el overhead de control sea mucho menos en redes de este tipo que en redes grandes, ya que por ejemplo un broadcast en esta red no tardaría mucho en propagarse como si en una red de cientos de dispositivos.

3.3. Escuela

Daremos ahora análisis a la capturas de la red de una escuela secundaria . Esta es una red mediana, comenzaremos analizando el tráfico por protocolo y la distinción entre Broadcast y unicast, luego procederemos al análisis de las fuentes S1 y S2 y sus respectivas propiedades descriptas por la Teoría de la información.

3.3.1. Tráfico por protocolo:

A continuación podemos ver los resultados de la medición de tráfico por tipo de protocolo:

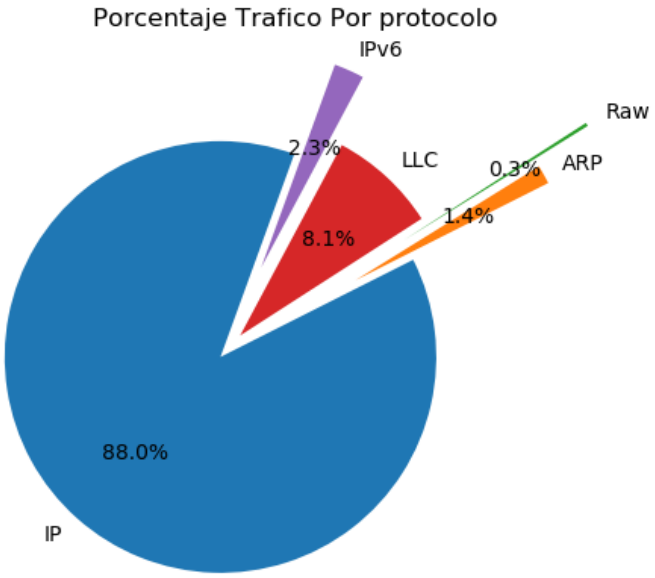

Figura 13: Tráfico por protocolo al de las capturas de la escuela

Se observan 5 tipos de protocolos distintos, 3 conocidos: `IP`,`IPv6` y `arp` y 2 desconocidos: `LLC` y `Raw`, Raw en realidad no es un protocolo sino que como dice el la información cruda adentro de un paquete Ethernet (u 802.3) sin protocolo.LLC es un protocolo utilizado para que pueden coexistir diferentes tipos de protocolos de red (ip,ipx,appletalk) y puedan ser transportados sobre la misma red.
El gráfico sólo muestra lo que hay adentro de cada trama Ethernet pero en el caso de esta red no todas las tramas de capa 2 eran Ethernet sino que tambien habían unos tipo de tramas llamadas 802.3 que son muy similares a las Ethernet. Se observa que la mayor parte de los paquetes son de los protocolos `IP` y `LLC` y en menor medida `IPv6` , los IP son esperables e ideales ya que suelen transpostar datos de usuario. no es el caso de los LLC ya que estos son de control y que sean tantos (8 % del total) da una clara limitación a las redes que quieran implementarlo, es decir que tolerar diferentes tipos de protocolos de red, tiene un costo alto.
Se observa que la cantidad de paquetes `ARP`, si bien no es despreciable no es tan abrumadora.

Se observan muchos mas paquetes de `IP` que de `IPv6`, lo que daria a entender que incluso en estas redes todavía el protocolo IPv6 no es algo desable de implementar a gran escala.

3.3.2. Tráfico Unicast-Broadcast:

Se muestra a continuación los porcentajes que indican el volúmen de paquetes de cada categoría (unicast o broadcast) independientemente de los protocolos usados. Para esto basta con ver el destino y analizar si es o no dirección de broadcast:

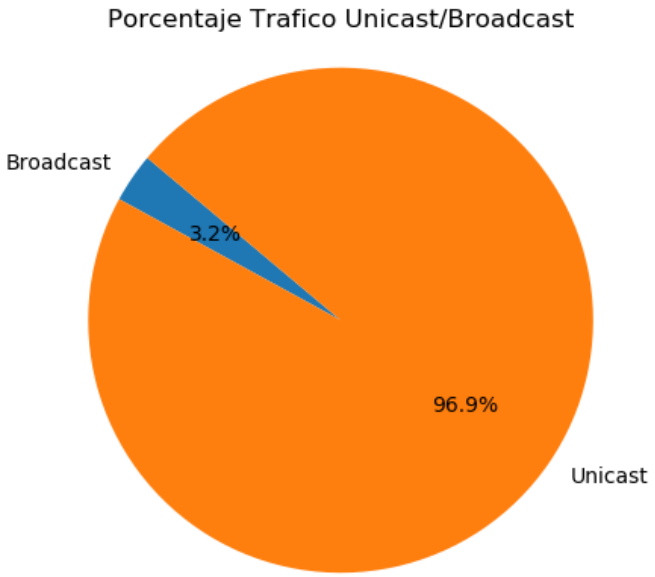

Figura 14: S1: escuela - Tráfico según tipo de destino de trama

Podemos notar según la Figura 14 que si bien los paquetes tipo **Unicast** predominan en comparación con los de tipo **Broadcast**, no es para nada despreciable la cantidad de paquetes de este último. Esto podría significar un sobrecargo importante en la eficiencia de la red por dos motivos, el primero es que los paquetes de Broadcast congestionan mas la red ya que deben ser enviados a todos los nodos de la misma, la segunda es que estos paquetes suelen ser de control por lo que serían netamente overhead.

3.3.3. Analisis de S1:

Analizaremos las propiedades de la fuente S1 pedida con las herramientas provistas por la teoria de la información. El gráfico a continuación mostrará la cantidad de información por símbolo comparándola con la entropía de la fuente y la entropía máxima teórica, hacemos especial distinción entre símbolos provenientes de distintos protocolos y símbolos con el mismo protocolo pero diferente categoría de destino(Broadcast-Unicast).

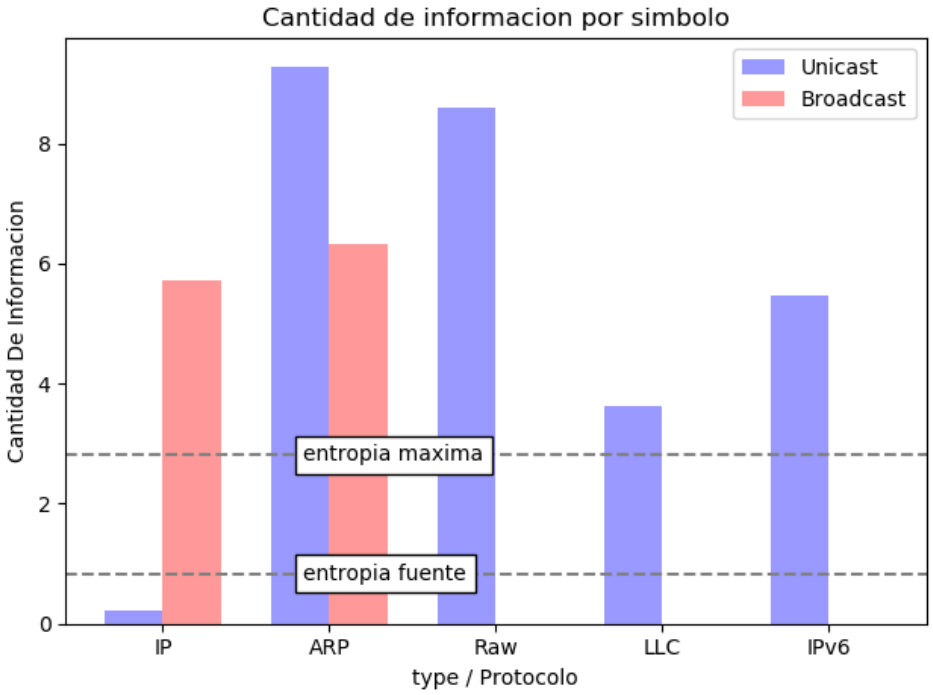

Figura 15: Escuela -Cant. información según símbolo distinguidos por Broadcast/Unicast

Se observa como todos los símbolos excepto $< IP, Unicast >$ están por arriba de la entropía y también por arriba de la entropía máxima, esto se debe a como observamos en los otros casos a que la cantidad (y por lo tanto la probabilidad) de los paquetes ip sea muy alta. Por lo anterior se hace muy fácil predecir el siguiente símbolo y por tanto baja la entropía de la fuente (como se ve en la figura la entropía esta muy por debajo de la entropía máxima). Se observa como es esperado que los símbolos de broadcast tengan información alta y de igual forma sucede con ARP. Esto es ideal porque habla de la probabilidad que aparezcan que deberia ser poca. No sucede de la misma forma con LLC, que esta cerca de la entropía máxima y por tanto tiene una frecuencia relativamente alta para lo esperado de un símbolo de control.

3.3.4. Análisis de S2:

Como fue explicado en la sección de metodologías nuestra fuente S2 planteada consta de las IPs de los diferentes nodos de la red. Estas fueron capturadas desde los paquetes ARP (como fue explicado en esa sección). En esta sección analizaremos las propiedades desde el punto de la teoria de la información de esta fuente generada.
Para el análisis de la información de fuente S2 graficaremos la información de los diferentes símbolos:

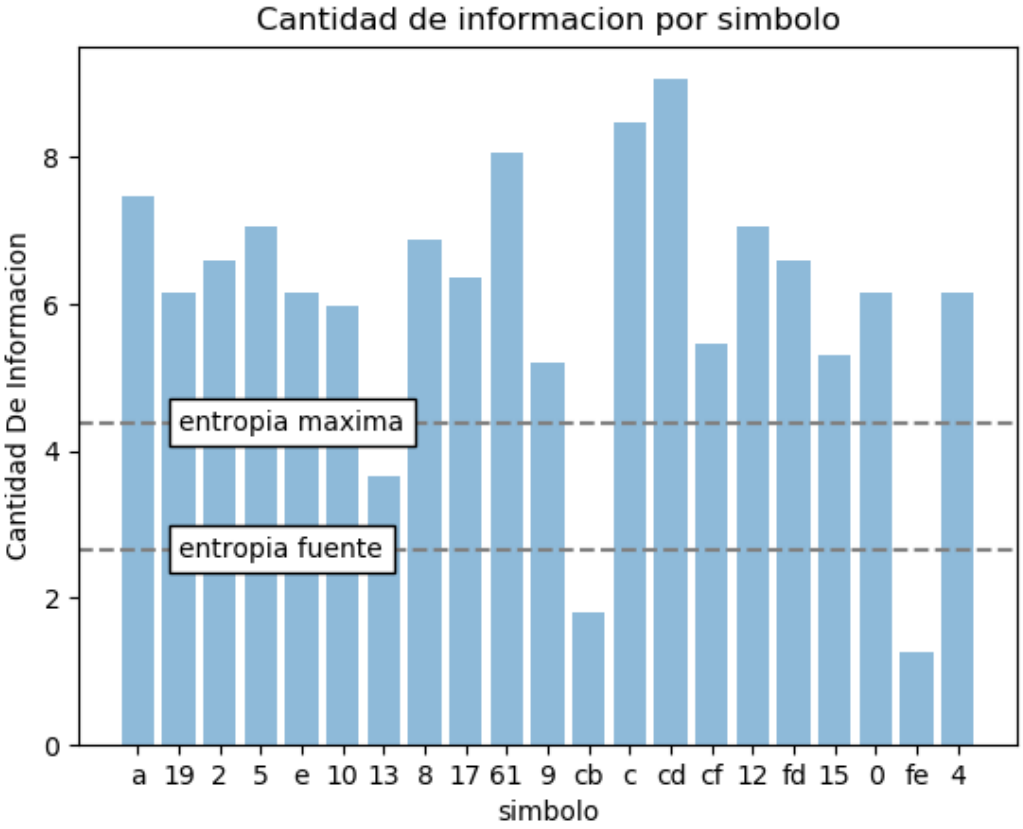

Figura 16: Escuela: de cada IP muestro solo su ultimo byte en hexa en la etiqueta

Se observa como la mayoria de los simbolos estan por arriba de la entropía incluso de la entropía máxima, esto nos dice que hay unos pocos símbolos, los que estan por debajo que se repiten con mucha frecuencia mientras que otros son mucho menos frecuentes. Es razonable pensar que aquellos símbolos cuya información es baja son los que tienen salida directa a internet y por lo tanto son mucho mas solicitados.
Se observa tambien que la entropía de la fuente no esta demasiado lejos de la máxima, aunque tampoco cerca, esto nos da una idea de que la capacidad de predecir un símbolo es moderada a baja, lo cual es ideal para transmitir la mayor cantidad de información posible.

3.3.5. Topología

Analizaremos la topología de la red, mediante un grafo de conexiones basado en los paquetes ARP. Planteamos un grafo dirigido donde se captura en los paquetes ARP las direcciones de origen y destino(el pedido en who-has) en los who-has y dibujamos un nodos por cada símbolo de S2 (cada IP) y un eje de origen a destino.

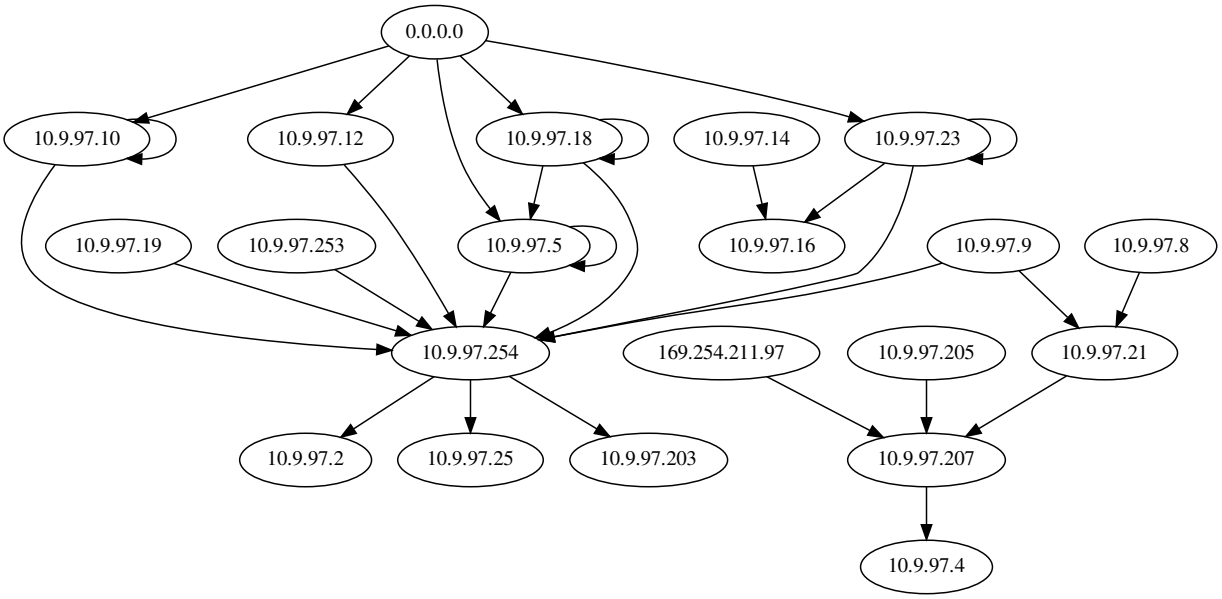

Figura 17: Grafo de red Escuela

Se observa de la Topología que la información fluye usualmente por el 10,9,97,254 desde los demás nodos, es esperable que entonces este símbolo tenga un nivel de información bajo y termine bajando la entropía de la fuente. Se observa el sentido "de abajo hacia arriba"particular que marca el flujo de la información.
Se observa como la red es de tamaño medio en comparación con las analizadas anteriormente.
Se ve también que hay algunas computadoras que no utilizan la salida mas común a internet sino que siguen caminos alternativos.

4. Conclusiones

Se concluye a partir de los experimentos planteados y las observaciones hechas:

1. Cuanto más grande la red, más overhead: el overhead tanto en cantidad de paquetes de Broadcast como protocolos de control (ARP,LLC) es mucho mayor. Es importante entonces cuando se realiza la construcción de una red prestar especial atención a que protocolos implementará esta y realizar esta pregunta en función del tamano de la misma.

2. Los protocolos de control como son LLC para poder utilizar múltiples direccionamientos son costosos, de nuevo sera necesario evaluar en cada caso si son necesario o no.

3. La baja en la entropía de una fuente en general se debe a un único o a unos pocos simbolos mientras que el resto tiene un nivel de informacion alto o acceptable, en general es un símbolo el que más desestabiliza la entropía de la fuente.

4. Los nodos de una red suelen llamar con frecuencia a las mismas IP esto es algo que hay que tener en mente a la hora de desarrollar algún nuevo protocolo. No tiene sentido suponer equiprobabilidad en los destinos de los paquetes.

5. Las redes por Wifi parecen ser mas desordenadas por los multiples access point y porque las redes cableadas fueron diseñadas de entrada mientras que las Wifi son redes que se van formando sobre la marcha y son mucho mas cambiantes. De todos modos a pesar de que esta es nuestra impresion, no tenemos evidencia suficiente para sostener fuertemente la conclusion ya que es preciso para eso evaluar redes Wifi y cableadas en las mismas condiciones.