DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Teoría de las comunicaciones

DC - UBA

Trabajo Práctico N°1

Integrante	LU	Correo electrónico
Rodrigo Kapobel	695/12	rok_35@live.com
Esteban Luciano Rey	657/10	estebanlucianorey@gmail.com
Nicolas Hernandez	122/13	nicoh22@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

1.1. Introduccion

Pueden destacarse dos protocolos importantes para lograr el correcto funcionamiento en la capa 2 y 3 del modelo OSI :

1.1.1. ARP

En red de computadoras, el protocolo de resolución de direcciones (ARP, del inglés Address Resolution Protocol) es un protocolo de comunicaciones de la capa de enlace, responsable de encontrar la dirección de hardware denominada MAC (Media Access Control) que corresponde a una determinada dirección IP. Para ello se envía un paquete (ARP request) a la dirección de difusión de la red (broadcast, MAC = FF:FF:FF:FF:FF:FF) que contiene la dirección IP por la que se pregunta, y se espera a que esa máquina (u otra) responda (ARP reply) con la dirección Ethernet que le corresponde.

Cada máquina mantiene una caché con las direcciones traducidas para reducir el retardo y la carga. ARP permite a la dirección de Internet ser independiente de la dirección Ethernet, pero esto solo funciona si todas las máquinas lo soportan.

En Ethernet, la capa de enlace trabaja con direcciones de hardware. El protocolo ARP se encarga de traducir las direcciones IP a direcciones MAC. Para realizar esta conversión, el nivel de enlace utiliza las tablas ARP, cada interfaz tiene tanto una dirección IP como una dirección física MAC.

ARP se utiliza en cuatro casos referentes a la comunicación entre dos hosts:

1. Cuando dos hosts están en la misma red y uno quiere enviar un paquete a otro.
2. Cuando dos hosts están sobre redes diferentes y deben usar un gateway o router para alcanzar otro host.
3. Cuando un router necesita enviar un paquete a un host a través de otro router.
4. Cuando un router necesita enviar un paquete a un host de la misma red.

1.1.2. IP

El protocolo de Internet (en inglés Internet Protocol o IP) es un protocolo de comunicación de datos digitales clasificado funcionalmente en la capa de red según el modelo internacional OSI.

Hay diferentes versiones de dirección IP actualmente. El más utilizado es IPv4. IPv6 es el sucesor propuesto de IPv4; poco a poco Internet está agotando las direcciones disponibles por lo que IPv6 utiliza direcciones de fuente y destino de 128 bits, muchas más direcciones que las que provee IPv4 con 32 bits.

2. Modelado del tráfico

El objetivo de este informe es realizar un análisis de carácter científico sobre el tráfico de red que se produce en la capa 2 y 3.

Se analizarán 3 (tres) redes distintas. Dos Wi-Fi y una cableada, de las cuales se obtuvieron 60000 muestras cada una.

Para procesar la información obtenida se modelarán las tramas de capa dos capturadas en las redes como fuentes de información de mememoria nula.

2.0.1. Fuente S_1

$S_1 = s_1, s_2, ..., s_q$, donde cada símbolo s_i está formado por la combinación entre el tipo de destino de la trama (unicast o broadcast) y el protocolo de la capa inmediata superior encapsulado en la misma. Por ejemplo, $s_i = <$ broadcast, ARP $>$.

Con esta fuente nos será sencillo observar porcentaje de tráfico broadcast sobre el tráfico total y mostrar el porcentaje de aparición de cada protocolo encontrado.

2.0.2. Fuente S_2

$S_2 = s_1, s_2, ..., s_q$, que tiene por objeto ayudar a *distinguir* los hosts de cada red.

Ya que el modelado de la fuente tiene que usar símbolos que se hallen hasta los paquetes de capa 3, se pueden utilizar las direcciones MAC, IP y los distintos mensajes ARP que se encuentren en los paquetes.

Un símbolo será *distinguido* cuando sobresalga del resto en términos de la información que provee. Con lo cual, al querer distinguir hosts de routers, buscamos símbolos que los evidencien mediante su baja probabilidad de aparición. Como dato a tener en cuenta, las redes capturadas poseen la característica en común de que son usadas por sus hosts para el acceso casi exclusivo a internet, siendo la comunicación local, casi despreciable. Esto nos permite inferir que la comunicación dentro de las mismas serán mayormente entre los host y el default-gateway, generando la hipótesis de que dentro de los IPs de la red, el del default-gateway será la más consultada por los dispositivos.

Dentro de los mensajes ARP distinguimos a las operaciones ”who-has” y ”is-at”, siendo la segunda la contestación de la primera. Tomando los ARP ”who-has”, según nuestra hipótesis, el destinatario que más veces se repita deberá ser el router default-gateway. No obstante, si uno de estos host esta consumiendo muchos datos (descarga de datos), distorsionaría la métrica.

Lo que se plantea entonces es utilizar todos los paquetes ARP y considerar un símbolo de la fuente como cada IP que aparezca en los paquetes (por cada paquete aparecen 2 símbolos). Esto quiere decir que si modelamos a cada dispositivo como un nodo de un grafo y las comunicaciones desde y hasta ese dispositivo como una arista, entonces estaríamos utilizando el grado del nodo como cantidad de repeticiones dentro de una muestra.

Como estamos asumiendo que el router va a poseer la IP de mayor presencia, le estaríamos dando mayor distinción a los host a nivel información.

3. Muestreo del tráfico

Como se ha mencionado, se realizaron 3 (tres) capturas de red. Dos para Wi-Fi (802.11) y una para Ethernet (802.3) de las cuales se tomaron 60 mil paquetes cada una.

1. Red Wi-Fi The App Master
2. Red Wi-Fi Hogareña Rivera 4370 CABA
3. Red Ethernet Hogareña Derqui

Se mostrarán los resultados obtenidos para cada una de las fuentes en cada red. Para reproducir los datos presentados, referirse al archivo README.md".

3.0.1. Red Wi-Fi The App Master

La captura se realizó el día 12 de Septiembre de 2017 en el horario que se llevaba a cabo la WWDC de Apple (Aproximadamente 14:20 hs). Al momento en la oficina había aproximadamente 7 personas, cada una con un celular y una notebook. La red consta de un Router que provee a los empleados de conexión a Internet.

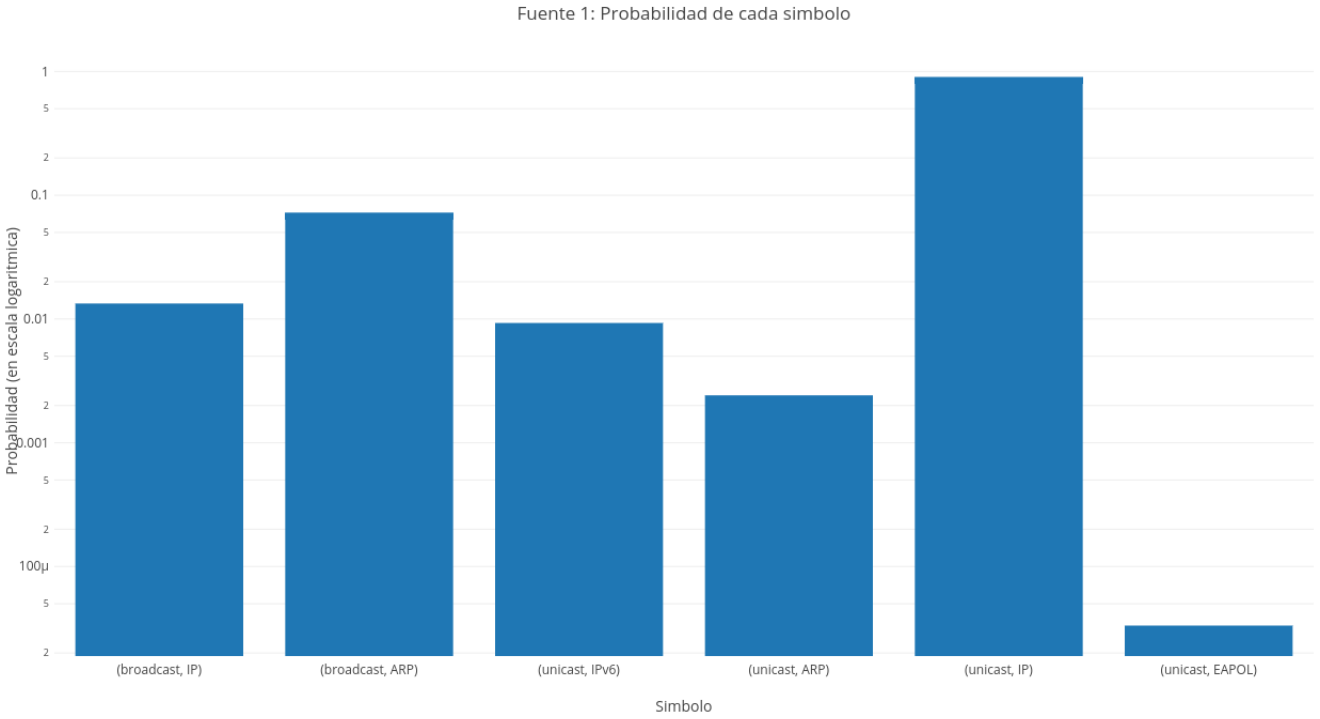

Figura 3.1: Probabilidad por símbolo fuente 1

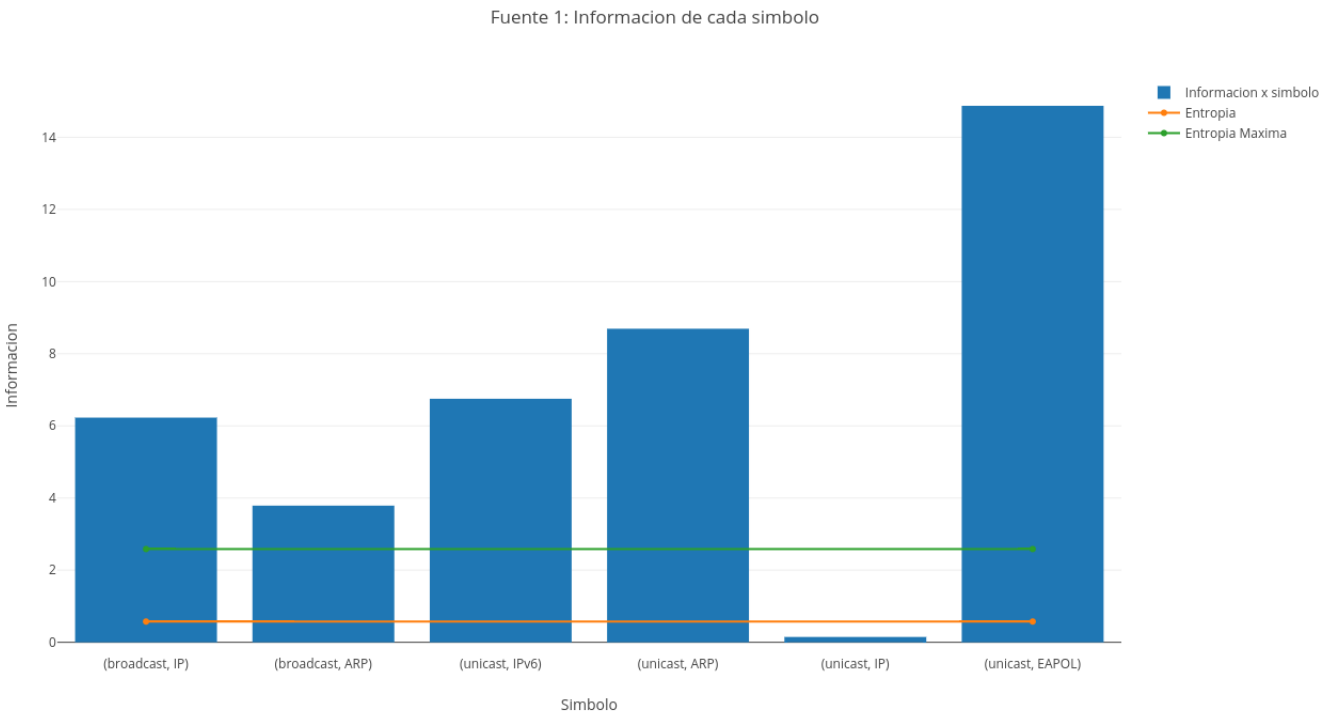

Figura 3.2: Información por símbolo fuente 1

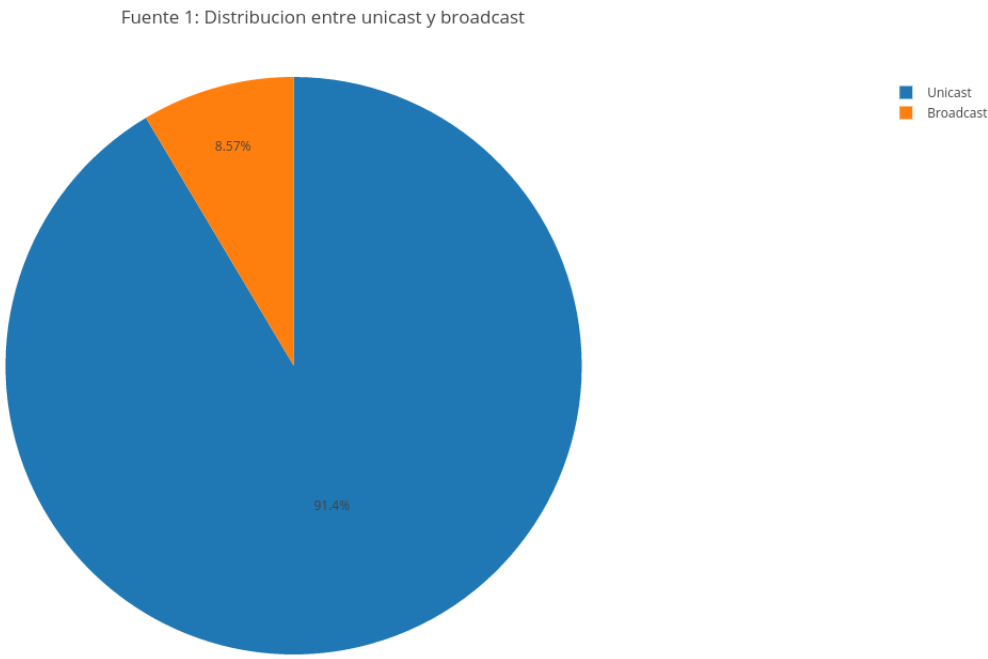

Figura 3.3: Porcentaje broadcast vs. unicast fuente 1

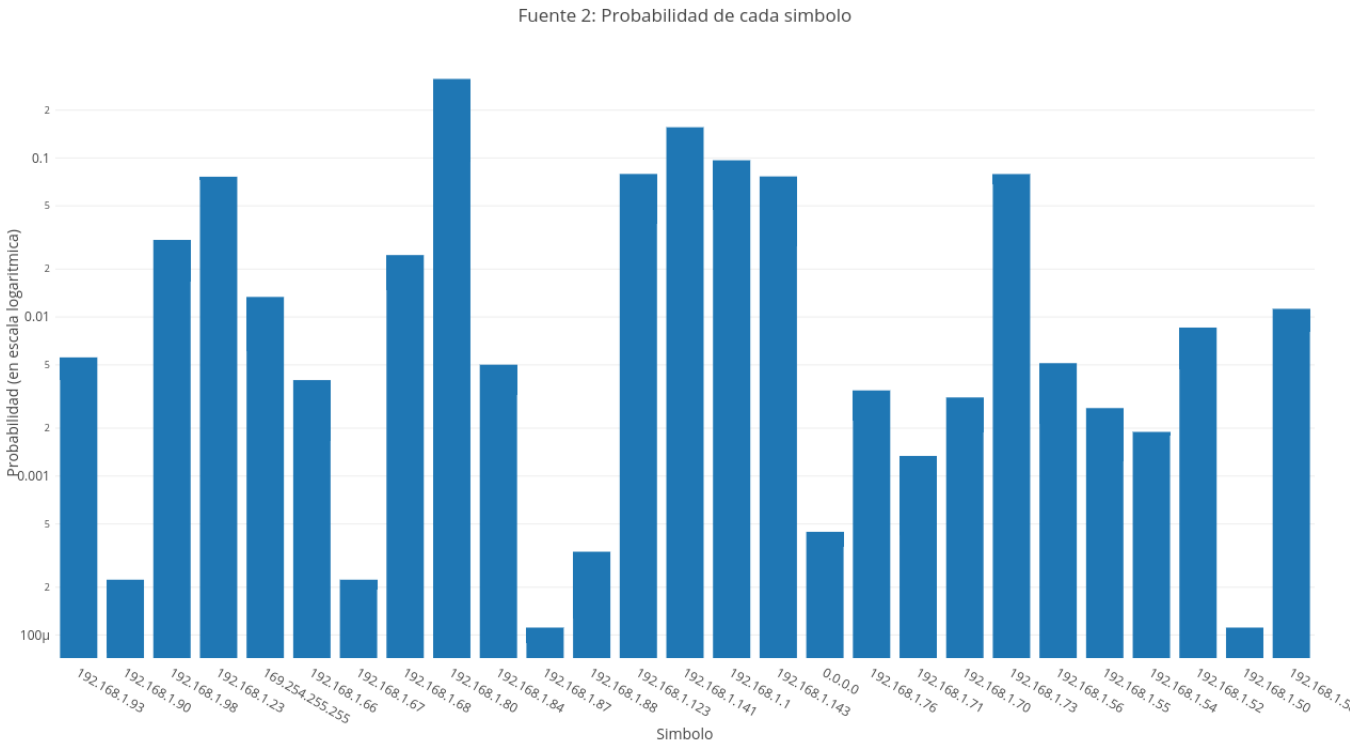

Figura 3.4: Probabilidad por símbolo fuente 2

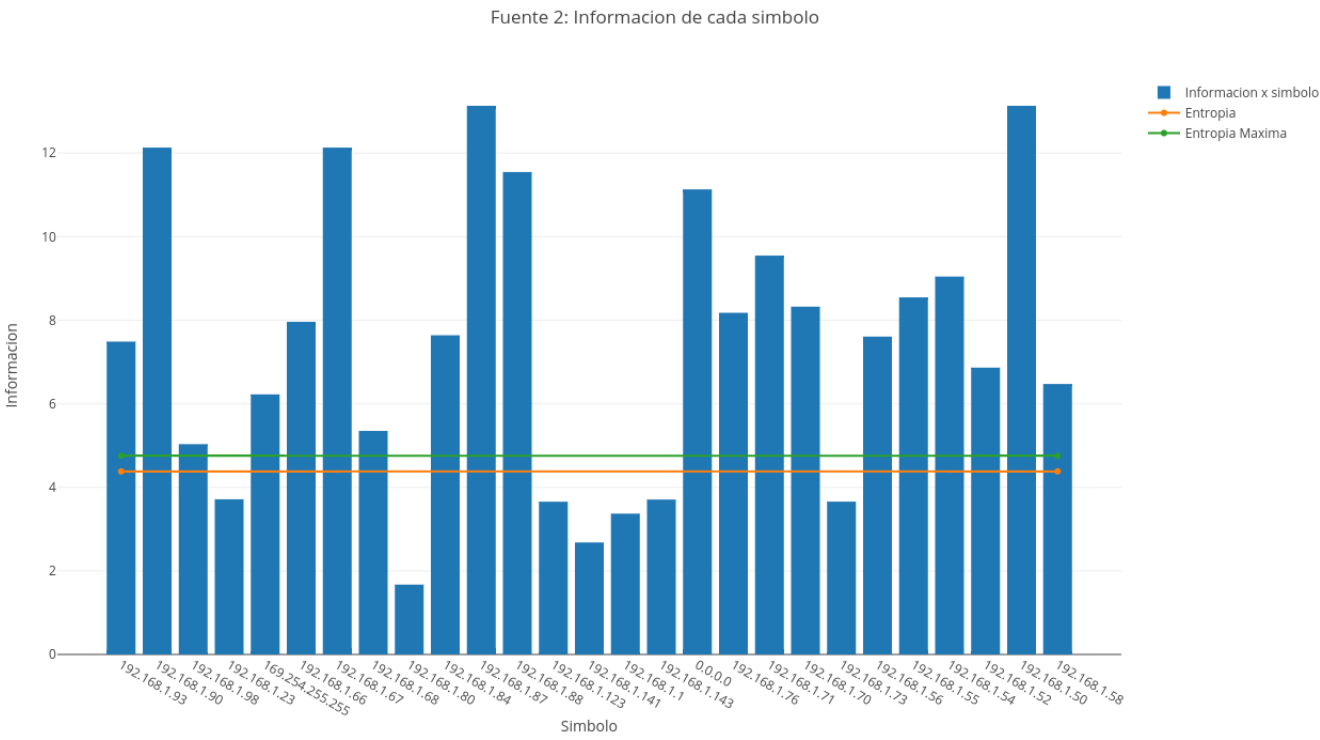

Figura 3.5: Información por símbolo fuente 2

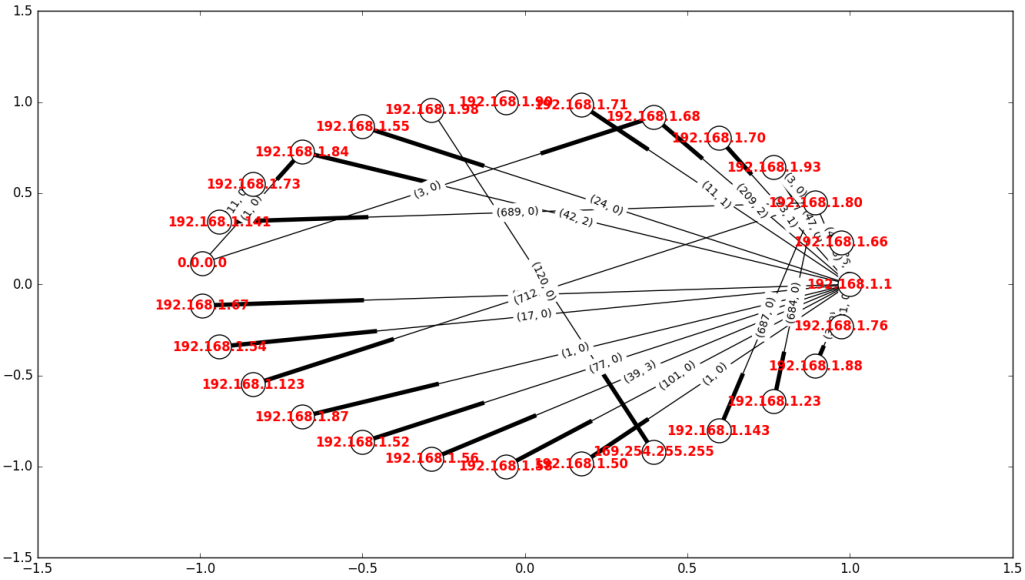

Figura 3.6: Información por símbolo fuente 2: Los nodos representan los distintos dispositivos de la red representados por IPs; las aristas simbolizan trafico de ida y vuelta, las etiquetas de las mismas son tuplas que representan el volumen de datos para un sentido y el otro: el sentido de mayor volumen esta representado por el sentido de la arista (la punta de la flecha vendria ser la seccion más ancha)

3.0.2. Red Wi-Fi Hogareña Rivera 4370 CABA

La captura se realizó el día 23 de Septiembre de 2017 aproximadamente a las 12:50 hs. Se desconoce la cantidad de personas exactas presentes al momento de la captura debido a que es una red compartida por los inquilinos de la casona. La red consta de un Router que provee a los inquilinos de conexión a Internet.

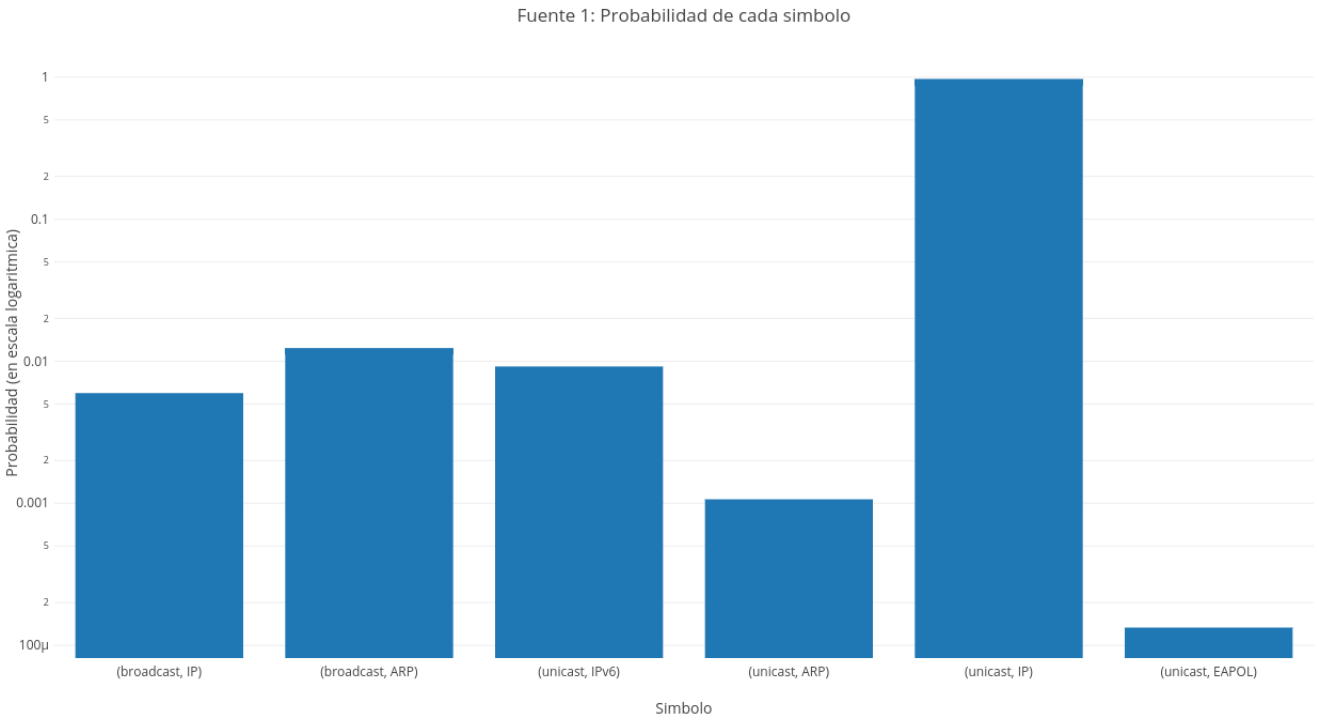

Figura 3.7: Probabilidad por símbolo fuente 1

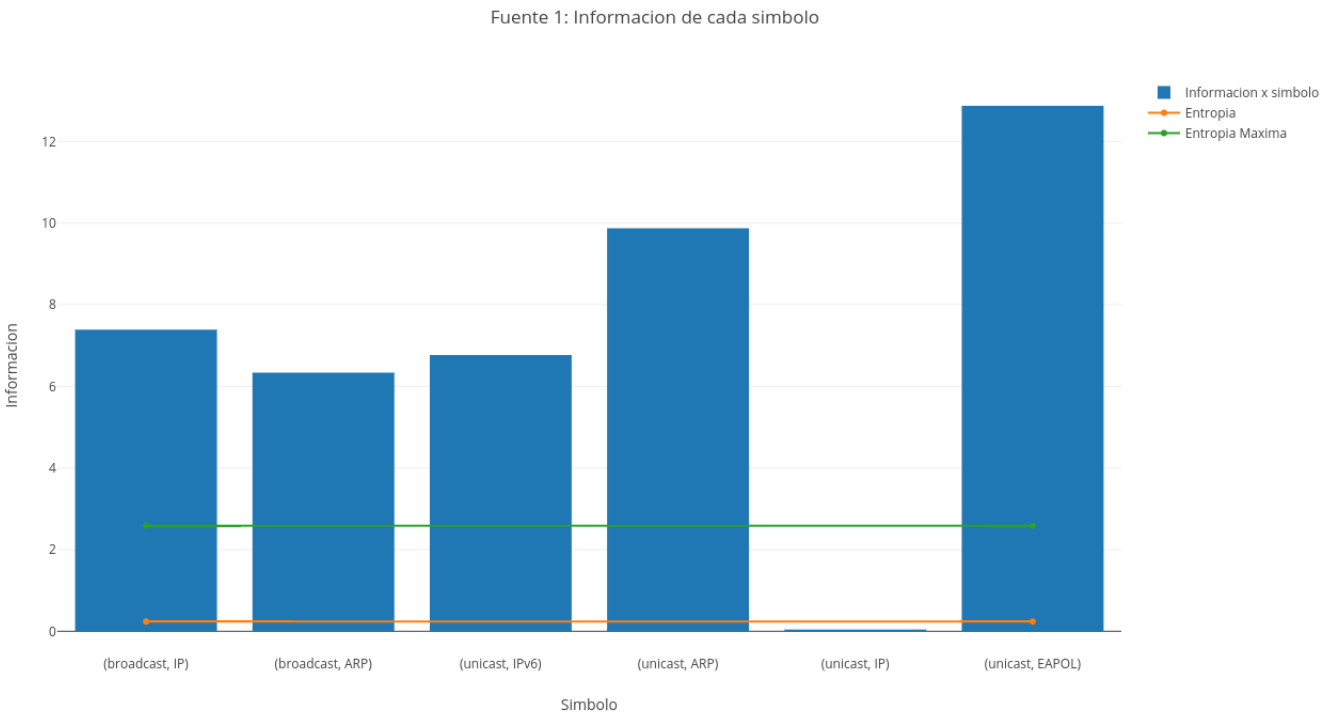

Figura 3.8: Información por símbolo fuente 1

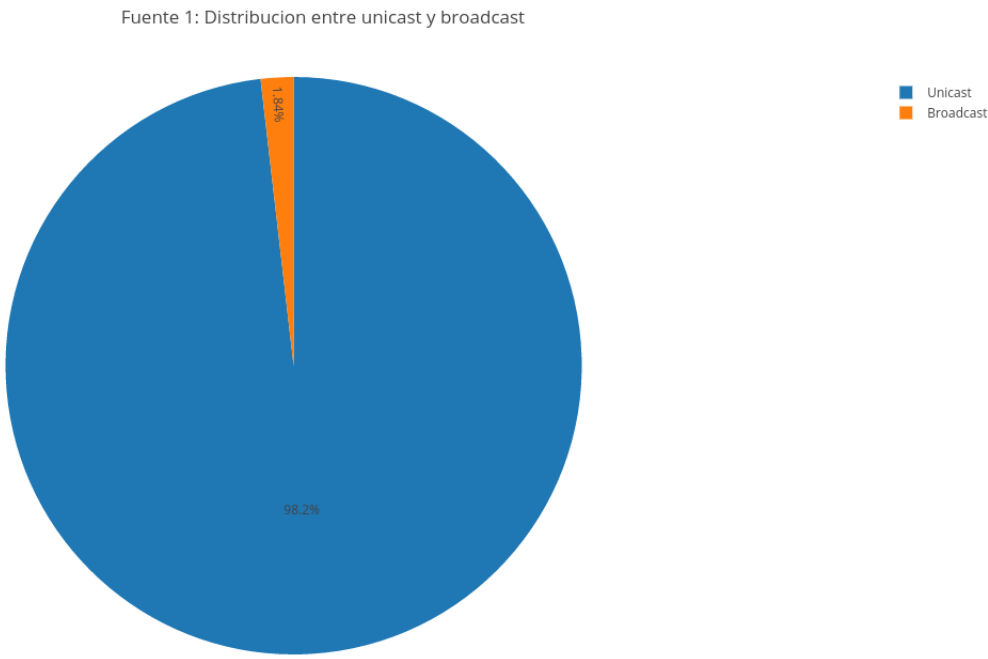

Figura 3.9: Porcentaje broadcast vs. unicast fuente 1

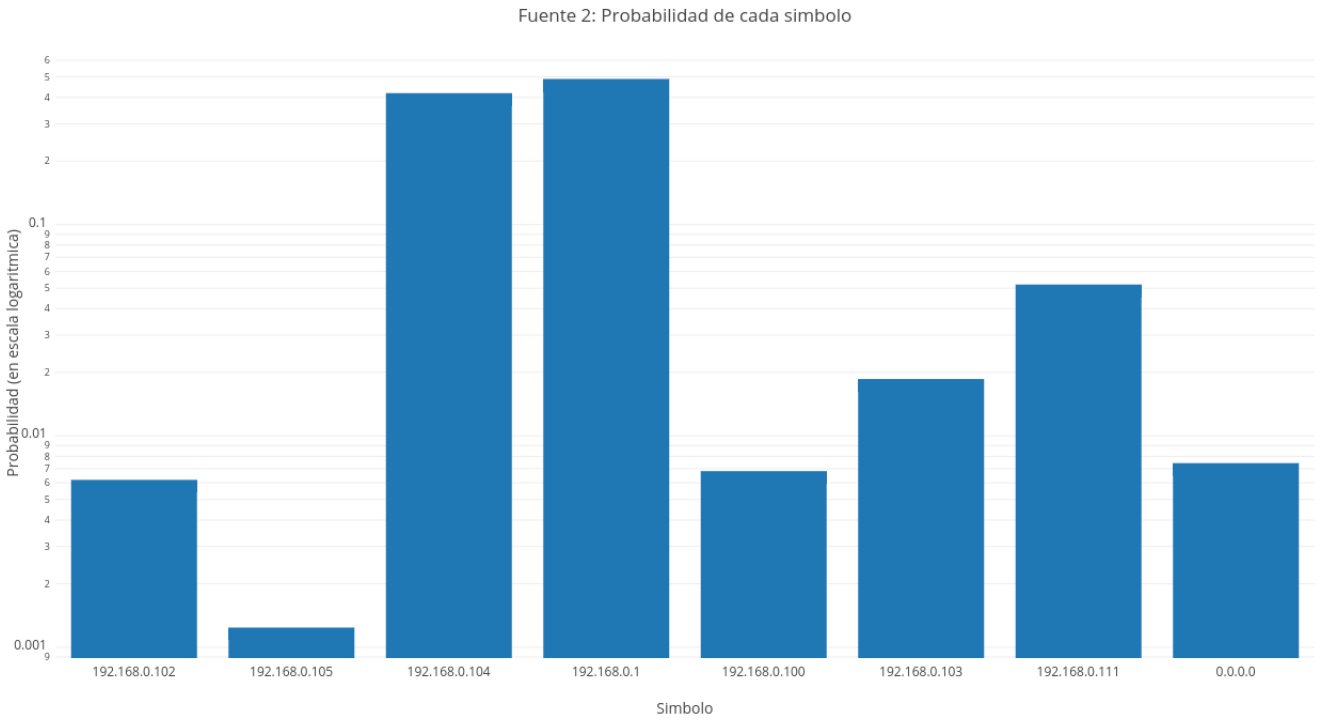

Figura 3.10: Probabilidad por símbolo fuente 2

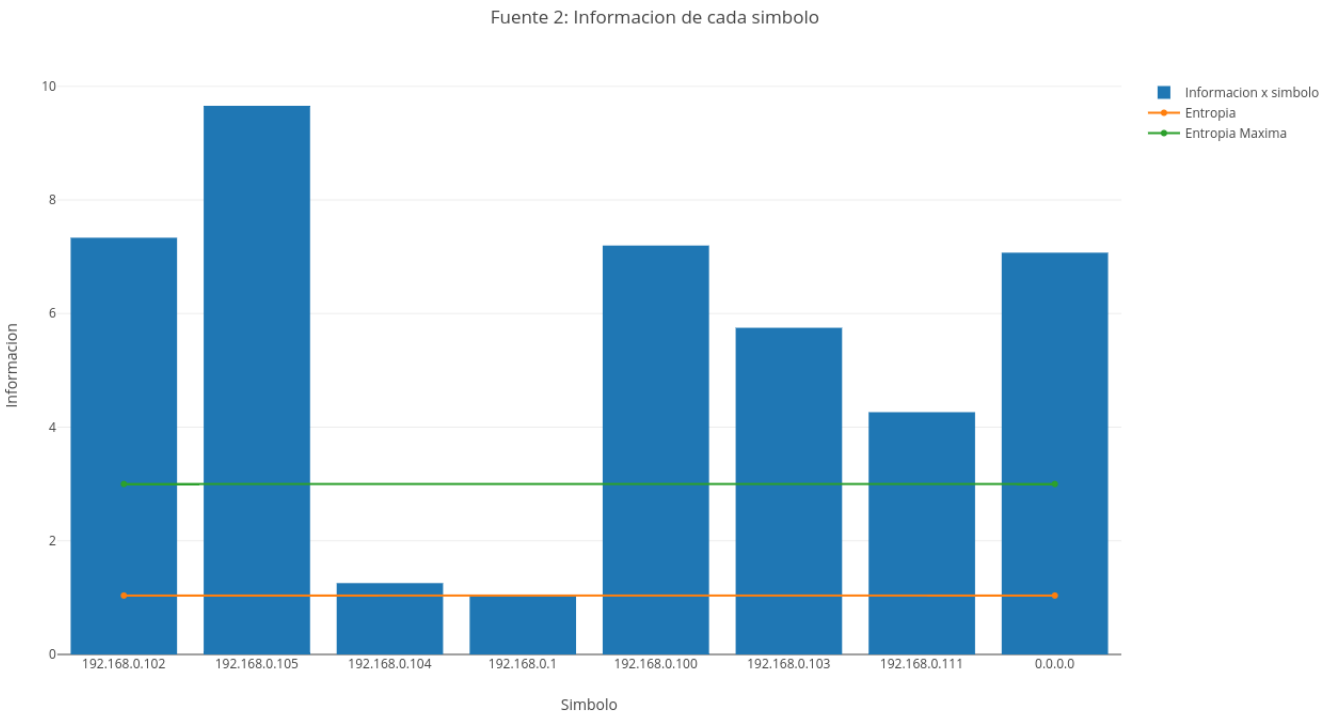

Figura 3.11: Información por símbolo fuente 2

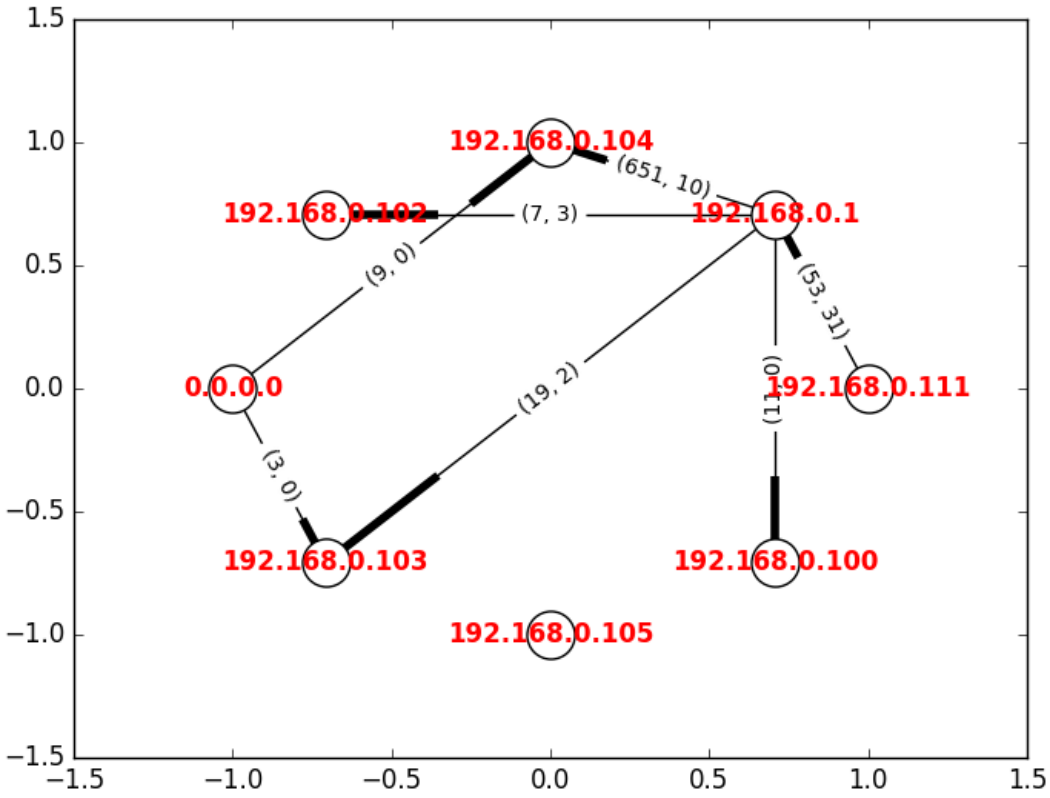

Figura 3.12: Información por símbolo fuente 2

3.0.3. Red Cableada Hogareña Derqui Provincia de Buenos Aires

La captura se realizó el 5 de Octubre de 2017 aproximadamente a las 11:00 hs. Durante la captura se encontraban aproximadamente 4 personas en el hogar con aproximadamente la misma cantidad de dispositivos electrónicos conectados a la red. El ordenador desde el cual se obtuvo la captura está conectado por Ethernet al router.

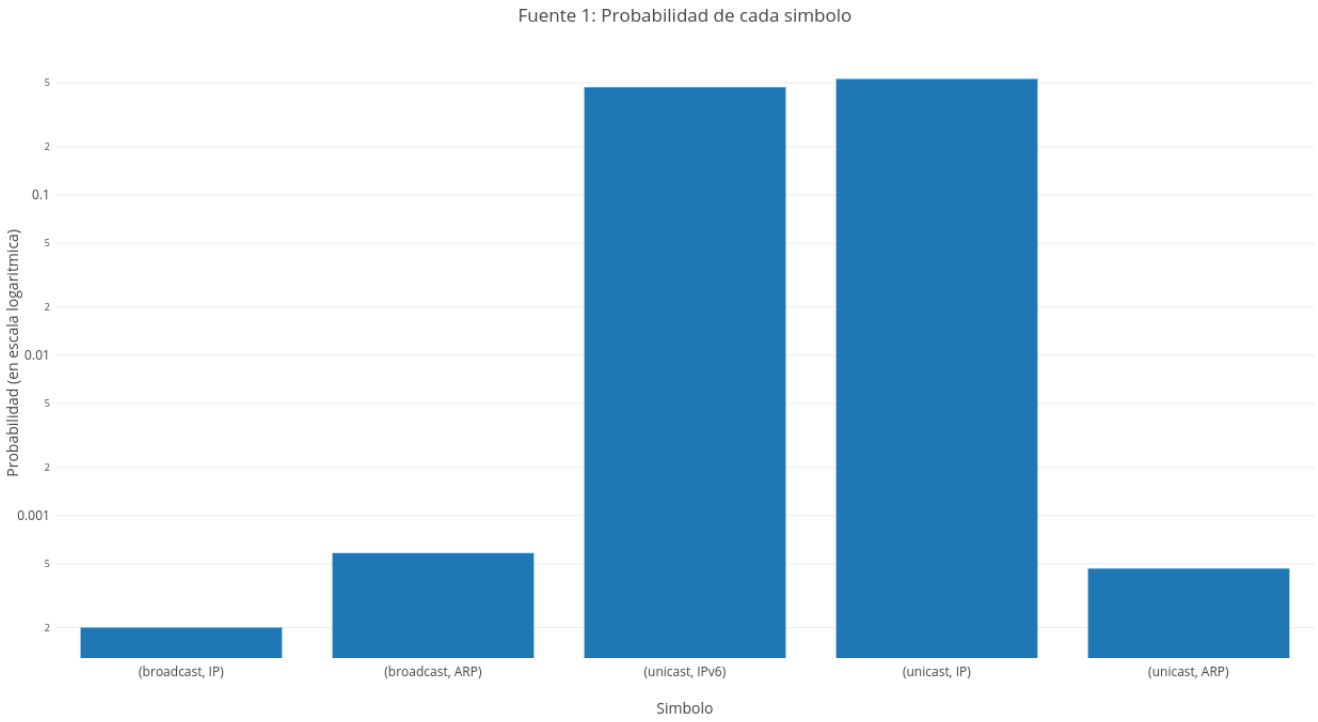

Figura 3.13: Probabilidad por símbolo fuente 1

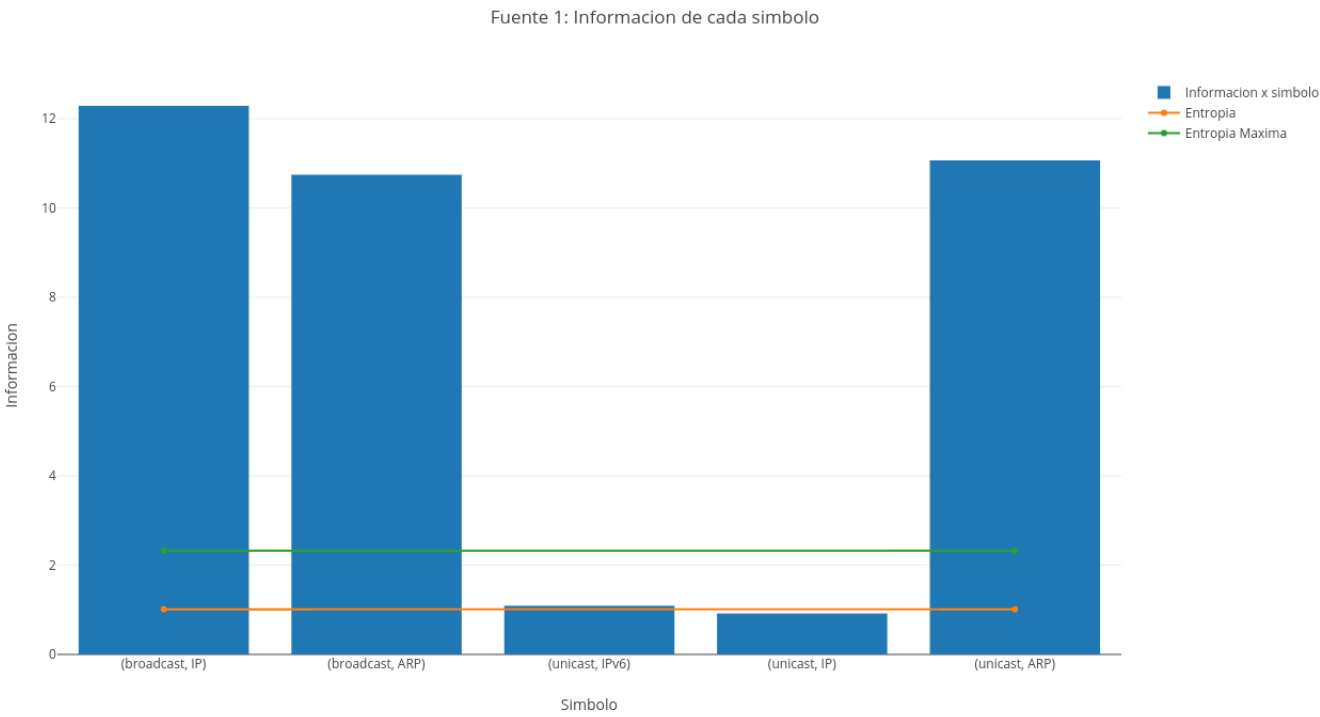

Figura 3.14: Información por símbolo fuente 1

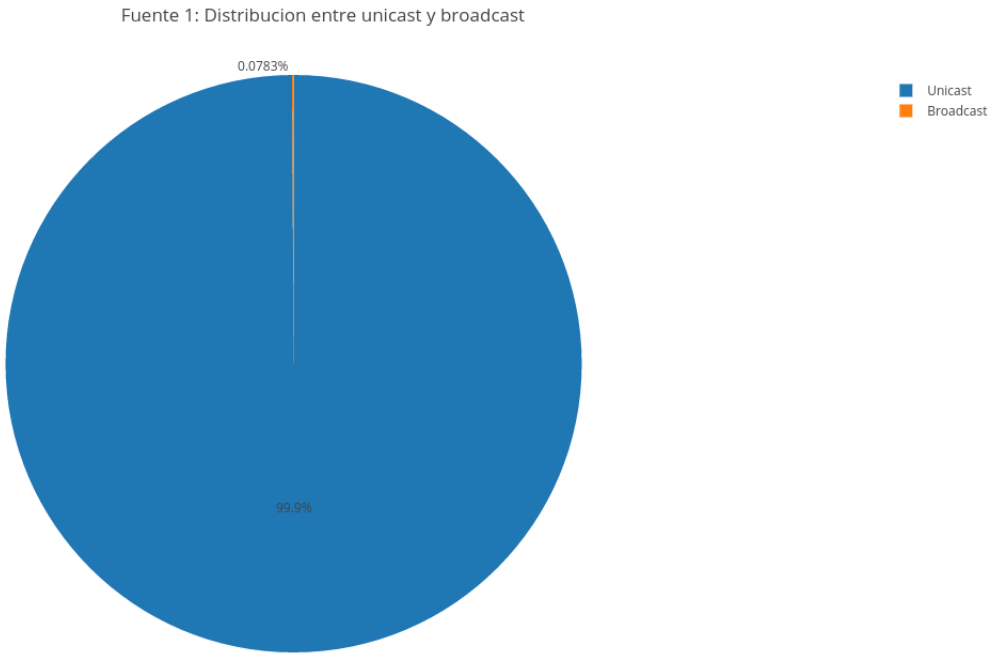

Figura 3.15: Porcentaje broadcast vs. unicast fuente 1

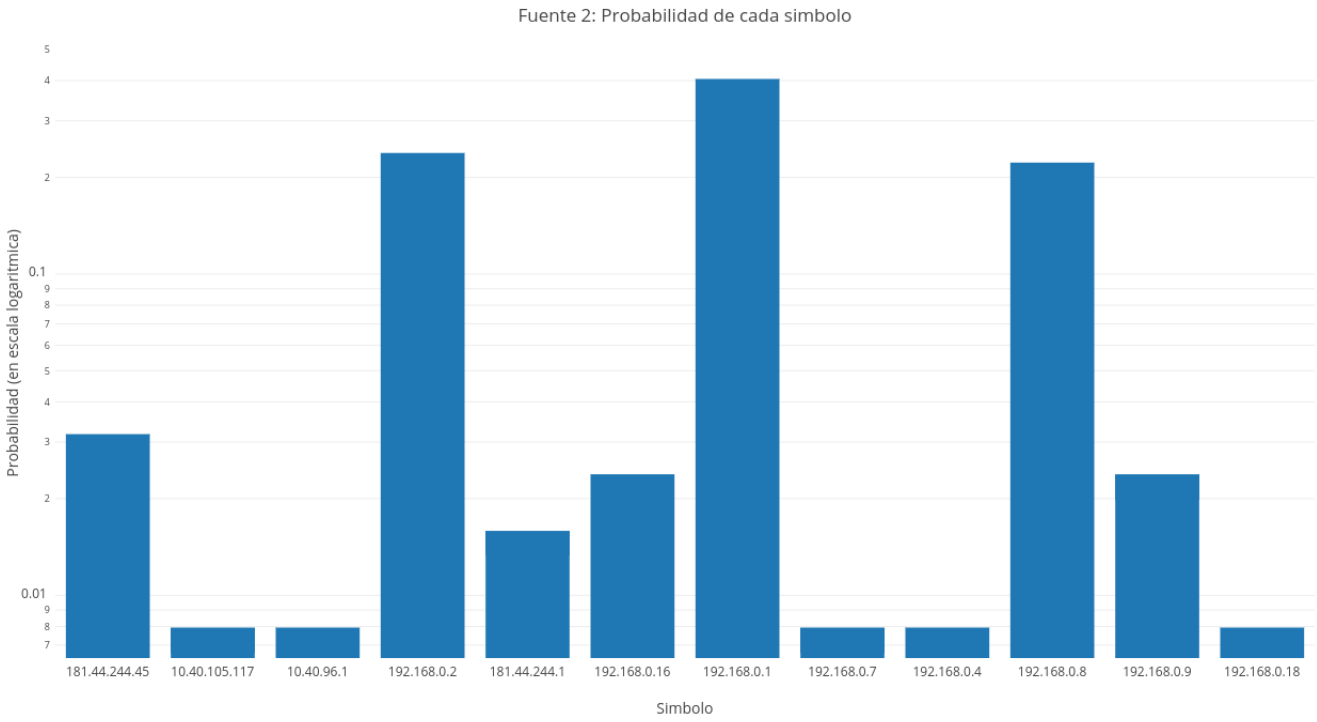

Figura 3.16: Probabilidad por símbolo fuente 2

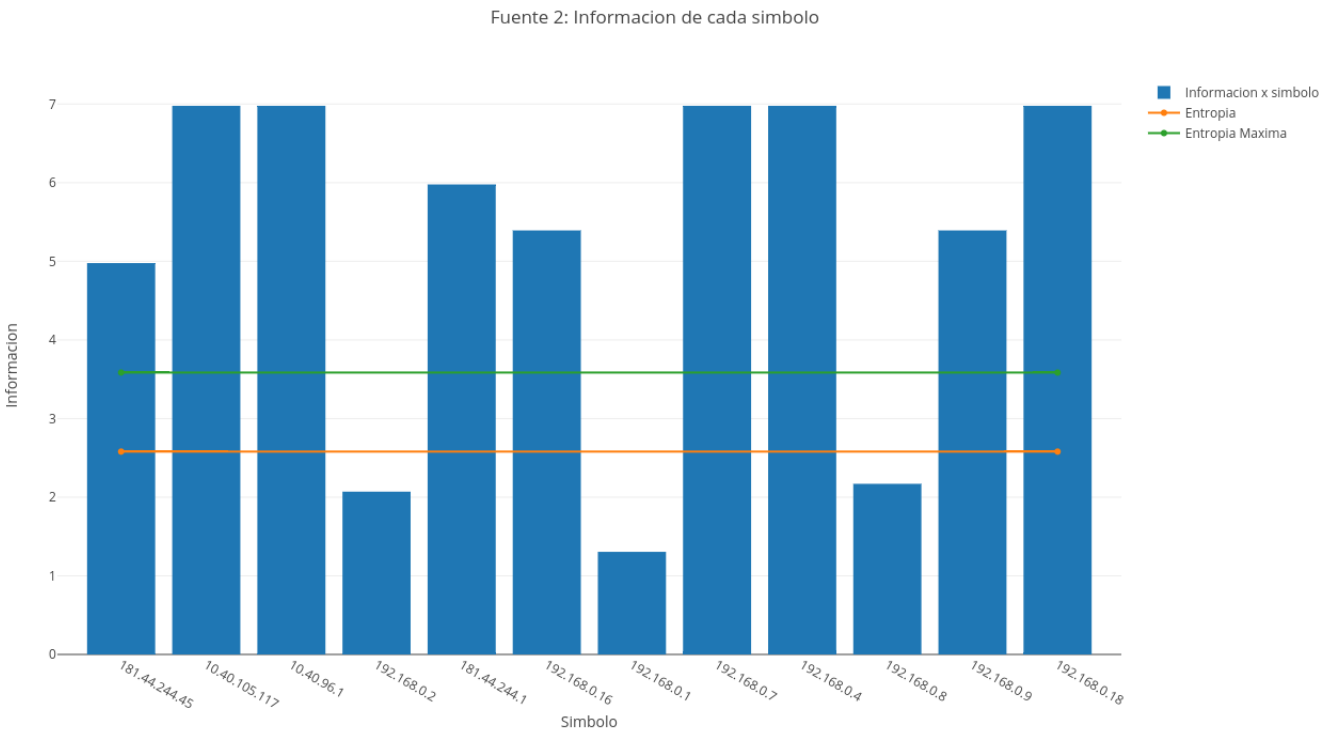

Figura 3.17: Información por símbolo fuente 2

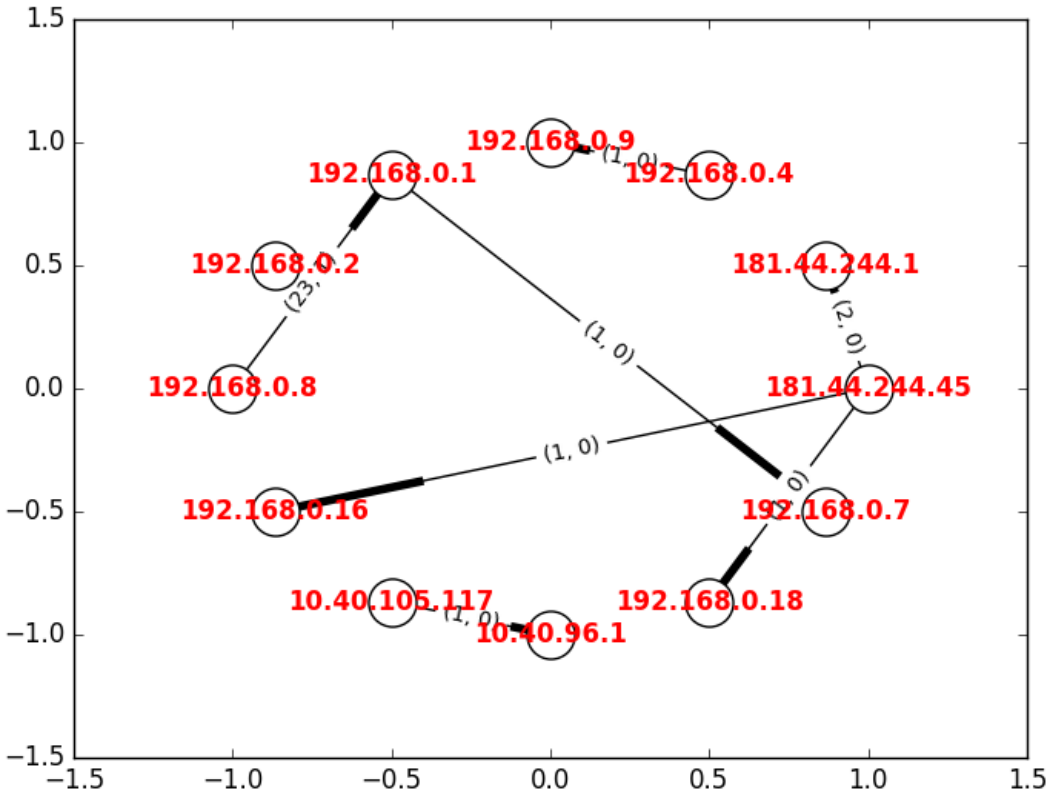

Figura 3.18: Información por símbolo fuente 2

4. Análisis de datos

4.1. Red Ethernet Hogareña Derqui

En esta red observamos que la cantidad del tráfico broadcast es muy pequeña (de hecho es la menor entre las 3 redes analizadas). Esta diferencia marcado entre tráfico broadcast y unicast influye en que la entropía de la fuente *S1* no se acerque mucho a la máxima. Los símbolos $IP - unicast$ y $IPv6 - unicast$ son muy frecuentes con respecto al resto de los símbolos, por lo tanto son los que menos información aportan.

Al haber tanta diferencia entre la información de los símbolos, la entropía es significativamente menor a la máxima. Para que la entropía sea máxima, deberíamos haber observado con la misma probabilidad todos los símbolos, para lograr eso hubiéramos necesitado más paquetes broadcast y más paquetes ARP.

En la fuente *S2* la entropía no es máxima, por lo tanto los paquetes ARP no aparecen equiprobablemente para los nodos de la red. Esto sugiere que hay nodos que reciben más respuestas *is-at*, o dicho de otra forma hay nodos que envían más paquetes que otros. Los símbolos distinguidos son 10.40.96.1, 192.168.0.7, 192.168.0.18 por tener una información muy por encima de la entropía y 192.168.0.1 por tener una información proporcionalmente chica. Coincidentemente, 192.168.0.1 es la dirección IP del router

en esta red. Como tiene poca información, significa que en la captura aparecieron una cantidad considerablemente mayor a la media de paquetes ARP que lo involucran. La razón de esto es que el router debe saber la ubicación de todos los nodos de la red para poder comunicarlos, entonces recibe muchos paquetes *is-at*.

En el mapa de red se observan IPs que no pertenecen a la red 192.168.0.0 . Investigamos usando la herramienta nmap y obtuvimos para uno de los hosts:

```
Nmap scan report for cpe-181-44-244-1.telecentro-reversos.com.ar (181.44.244.1)
Host is up (0.021s latency).
Not shown: 995 closed ports
PORT     STATE SERVICE
2001/tcp open  dc
2126/tcp open  pktcable-cops
4001/tcp open  newoak
6001/tcp open  X11:1
9001/tcp open  tor-orport
```

Lo cual nos da el indicio de que estas direcciones pertenecen a algún router del ISP de la red. También escaneamos la dirección 10.40.105.117 y nos dió que su puerto HTTP estaba abierto. Al acceder con un navegador web nos lleva a una pagina de *login* similar a la del router de la red.

4.2. Red Wi-Fi Hogareña Rivera 4370 CABA

En esta red vemos que la proporción de tráfico broadcast es considerablemente mayor a la observada en el red anterior. La entropía en la fuente *S1* es muy baja, esto se debe a que el símbolo $IP - unicast$ tiene una información muy baja y los paquetes de este tipo constituyen la mayor parte del tráfico de la red. El símbolo $EAPOL - unicast$ es distinguido al ser el de mayor información.

En esta red encontramos un protocolo inesperado: EAPOL. Este protocolo usado en IEEE 802.1X, cuyas siglas significan *Extensible Authentication Protocol over LAN*, es de control de acceso, tiene como objeto proteger recursos de red y requerir autenticación de aquellos clientes que deseen accederlo dentro de una LAN. Es usado por los sistemas de autenticación WEP, WPA y WPA2.

En la fuente *S2* la entropía tampoco es máxima. Los nodos que se distinguen en este caso son 192.168.0.102, 192.168.0.105 por su gran cantidad de información y 192.168.0.1 por estar por debajo de la entropía. Como se puede observar en el mapa de red, el nodo con más ejes es 192.168.0.1, el cual es el router y el símbolo con menor información. Además el host desde donde se realiza la captura (192.168.0.104) también presenta información baja.

4.3. Red Wi-Fi The App Master

Esta red posee el mayor porcentaje de mensajes broadcast de las redes sensadas, llegando casi a tener el 10 % del total. Se puede observar que aunque el broadcast del protocolo IP tiene baja probabilidad de ocurrir (similar al unicast de IPv6), el de ARP es el responsable del peso de los mensajes totales. Si comparamos las proporciones con las otras 2 redes, podemos observar que estos mensajes ARP son mucho más elevados, pudiendolo atribuir a la diferencia de cantidad de hosts con las otras muestras: a más dispositivos, más mensajes de sincronización necesarios.

Nuevamente se detecta el protocolo EAPOL, siendo uno de los símbolos más distinguidos de nuestra

fuente. El símbolo que menos se distingue es el de unicast del protocolo IP, usado para el transporte de datos. Los símbolos que representan a los protocolos de control ARP muestran que los $who-has$ enviados por broadcast son mucho más que los $is-at$ enviados por unicast.

Modelando con la fuente 2 podemos distinguir 2 grupos de dispositivos: aquellos con mayor información que la entropía máxima de la fuente, y los de menor información. A sabiendas que el $default-gateway$ se encuentra en la IP 192.168.1.1, podemos ver en este caso, que la fuente 2 no nos dejaría muy en claro su distinción de los demás dispositivos pertenecientes al segundo grupo. Por el otro lado podemos ver que la mayoría de los hosts tienen en promedio mayor información que la entropía de la fuente.

Dentro de los nodos más distinguidos podemos ver a las IPs 192.168.1.87, 192.168.1.50 y 192.168.1.90, como aquellos que menor actividad reportan y el 192.168.1.141, 192.168.1.123 y 192.168.1.23 como los de mayor actividad, pudiendo indicar hosts que estaban realizando bajada de archivos desde internet, o streaming.

5. Conclusiones

Mediante 3 redes distintas, se pudo observar diferencias relacionadas al modo de acceso al medio de información y la dimensión de las mismas. Entre un enlace por Wi-Fi y uno cableado se encontró la diferencia de la presencia del protocolo EAPOL en la primera, y su ausencia en la segunda.

En consideración a la distinción entre paquetes enviados a broadcast y unicast, se pudo observar que la proporción del primero tiende a ser mucho menor a la del segundo (casi 9 veces menor); pero que al aumentar la cantidad de dispositivos en la red, la diferencia empieza a disminuir. Esto es atribuible a la mera pertenencia de los dispositivos en la red (osea la necesidad de usar los mecanismos de control) y la improbabilidad que cada uno de estos dispositivos este consumiendo la misma cantidad de datos (es decir que no por estar en la red, estén consumiendo datos).

Para las fuentes S1 y S2, los símbolos no son independientes. Si se envía un paquete ARP $who-has$ esto aumenta la probabilidad de que aparezca un paquete $is-at$. Incluso el tráfico IP aumenta la probabilidad de recibir una respuesta. Este tipo de situaciones causa que la entropía de la fuente disminuya.

En base a lo concluído, surge preguntar que podría ocurrir en una red wifi pública de acceso masivo (BA WiFi por ejemplo) en cuanto a distribución entre paquetes broadcast y unicast. Otro experimento pendiente es el de comparar la cantidad de mensajes is-at frente a who-has que se producen en una red en donde prime la conexión entre hosts, como por ejemplo un torneo de LAN (lugar óptimo de sniffing un ciber). Para esta ultima red, sería interesante ver si el modelado en S2 permite detectar la maquina servidora del juego.