DEPARTAMENTO
DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 1

Wiretapping

23 de septiembre de 2023 — Teoría de las Comunicaciones

Integrante	LU	Correo electrónico
Santesteban, Martín	397/20	martin.p.santesteban@gmail.com
Schwartzmann, Alejandro Ezequiel	390/20	a.schwartzmann@hotmail.com
Stemberg, Uriel Nicolás	213/20	uri.stemberg@gmail.com
Sztajn, Martin	181/20	martinsztajn@gmail.com

Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria - (Pabellón I)
Intendente Güiraldes 2610 - C1428EGA
Ciudad Autónoma de Buenos Aires - Rep. Argentina
Tel/Fax: (++54 +11) 4576-3300
http://www.exactas.uba.ar

Índice

1. Introducción

En este trabajo buscamos entender el funcionamiento de una red, y además lograr analizar y diagnosticar su estado y rendimiento. A través del análisis de protocolos de red, mediante la captura y estudio de los paquetes que circulan en una, buscamos obtener información valiosa sobre su funcionamiento y características.

Para llevar a cabo este análisis, es fundamental introducir algunos conceptos clave de la teoría de la información:

- **Fuente**: En nuestro contexto, la fuente es la red misma desde donde se originan los paquetes que estamos analizando.
- **Símbolo**: En este contexto, un símbolo es una representación única de un paquete específico que circula en la red, definido el tipo de destino (unicast o broadcast) y el protocolo utilizado, definido por el campo **type** del frame Ethernet.
- **Entropía**: Es la medida de la incertidumbre o sorpresa de una fuente de información. En nuestro estudio, calculamos la entropía para entender la variabilidad y la previsibilidad de los símbolos (paquetes) en la red. La entropía (H) se calcula:

  $$H(X) = -\sum p(x) \log_2 p(x)$$

  donde $H(X)$ es la entropía de la variable aleatoria X, $p(x)$ es la probabilidad de cada símbolo x y la sumatoria se realiza sobre todos los símbolos posibles.
- **Información**: La información de un símbolo específico se refiere a cuánto nos sorprende ese símbolo cuando lo observamos. Por lo tanto, un símbolo con menos frecuencia proporcionará más información que uno que aparece con más. La cantidad de información (I) asociada a un evento x se calcula como:

  $$I(x) = -\log_2 p(x)$$

  donde $I(x)$ es la cantidad de información del símbolo x y $p(x)$ es la probabilidad de ocurrencia del símbolo x, que en nuestro estudio se asoció a la razón entre las apariciones de las tuplas simbolo y el total de paquetes recolectados.

2. Métodos y Condiciones de la Experimentación

Respecto al código realizado, ayudándonos del provisto por el enunciado, implementamos un sniffer de paquetes utilizando la biblioteca Scapy de Python. Este script permite capturar paquetes que circulan en la red y clasificarlos según su tipo de destino (Unicast o Broadcast) y su protocolo (IP, ARP, IPv6, entre otros).

Para nuestro experimento filtramos los paquetes y solamente almacenamos en un diccionario los que tenían una capa Ethernet, y limitamos la cantidad, mayor a 10k pero menor a 30k paquetes. Luego, a cada paquete lo clasificamos según su protocolo y destino.

Posteriormente, calculamos la entropía de la red y la información provista de cada símbolo de la fuente. Definimos un símbolo como una tupla <destino, protocolo> (ejemplo <Unicast, IP>).

2.1. Descripcion de las redes

Para el experimento decidimos analizar cuatro redes en contextos distintos:

1. Red Hogareña en minima cantidad de uso
2. Red Hogareña en uso habitual
3. Facultad en horario pico
4. Facultad en horario poco concurrida

El sniffer fue ejecutado en tres redes distintas para obtener una muestra, cuatro veces en total. A continuación detallamos las condiciones específicas:

N°	Red	Ubicación	Fecha, Hora	Muestra
1	ClaroFibra24G8590	Florida, Vicente Lopez	14 de septiembre, 03hs a 11hs	25k paquetes
2	ClaroFibra24G8590	Florida, Vicente Lopez	13 de septiembre, 14hs	25k paquetes
3	UBA-WIFI Pabellón 1	UBA, Pabellón 1	13 de septiembre, 17 hs	12.5k paquetes
4	UBA-WIFI Pabellón 1	UBA, Pabellón 1	13 de septiembre, 20hs	12.5k paquetes

Cuadro 1: Detalles de las ejecuciones de la captura de paquetes

Es importante aclarar que todas las redes utilizaban teconologia IEEE 802.11n.

2.2. Modelo de Nodos distinguidos

Proponemos un modelo donde la fuente que tomamos es $S_2 = \{s_1, ..., s_n\}$ donde cada s_i representa la IP destino de un paquete ARP.

El criterio de identificación de los símbolos será su entropía, la cual debe ser menor a la entropía de la fuente, específicamente buscamos encontrar las IP que más paquetes reciben de la red, las cuales asumimos que serán el router y los dispositivos que más tráfico demandan. Llmaremos nodo distinguido a la dirección IP que mas paquetes ARP UNICAST recibio, que será el simbolo cuya aparición menos información proveera.

Para implementar este modelo, ejecutamos el sniffer en la red pública **UBA WIFI Evento**, por ser la de mayor tamaño y tráfico, y luego filtramos los paquetes por su capa ARP y almacenando la IP destino, luego calculamos la entropía de la fuente y la información provista por cada símbolo y los clasificamos por dicha información. La captura de 100 paquetes se llevó a cabo el dia 20 de septiembre, entre las 17:30 y las 18:45 horas.

3. Resultados de los experimentos

Presentamos a continuación, por cada red analizada, una exposición de las muestras tomadas junto con un analisis de las métricas. Adicionalmente se expondran los resultados del modelo de nodos distinguidos.

3.1. Red ClaroFibra24G8590 - 3am a 11 am

Entre los 25.000 paquetes capturados, se presentó el siguiente alfabeto:

$$A = \{(U, IP), (U, IPv6), (U, 35130), (U, ARP), (U, PNAC), (B, IP), (B, ARP)\}$$

donde U := UNICAST y B := BROADCAST. En la figura 1 se expone la proporción de paquetes por protocolo, quedando a la vista que no se trata de una fuente de simbolos equiprobables. Entre los protocolos habituales, se capturaron ademas paquetes con protocolo PNAC (Port-Based Network Access Control) y 35130, un protocolo que no se encuentra registrado en los estandares IEEE 802.

La figura 2 expone la cantidad de paquetes con destino UNICAST y BROADCAST en función al protocolo, dejando contemplar que solo los protocolos IPv4 y ARP presentaron paquetes broadcasteados. Además, el 93.672 % (23418) de los paquetes fueron UNICAST mientras que el 6.328 % (1582) restante fueron BROADCAST. Esto era esperable dada la pequeña LAN en la que se encuentra la computadora donde se corrió el experimento, que en efecto, es de una casa.

Por último, la figura 3 muestra la cantidad de información de cada símbolo. El valor teórico máximo de entropía es 2.807 y se observó una entropía igual a 2.04793, siendo la medición que más se acercó al valor teórico.

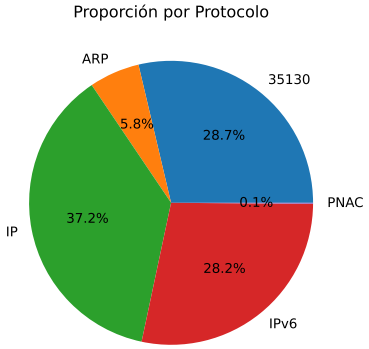

Figura 1: Proporción de paquetes por protocolo.

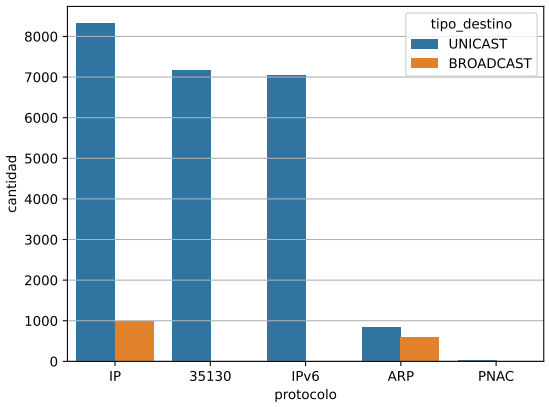

Figura 2: Proporción de paquetes de broadcast y unicast con respecto a los protocolos.

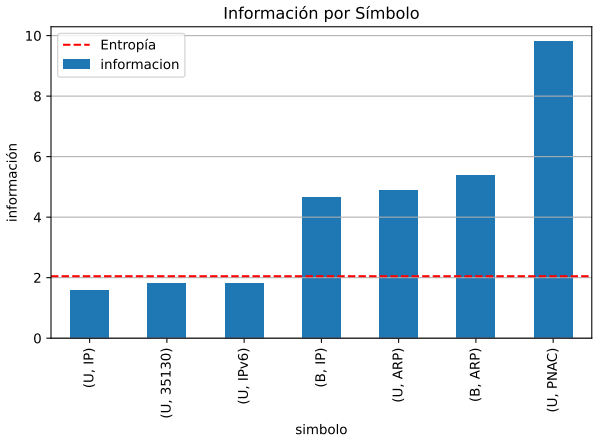

Figura 3: Cantidad de información por símbolo comparado con el valor de la entropía, donde U:= UNICAST y B := BROADCAST.

3.2. Red ClaroFibra24G8590 - 14hs

Se hizó otra medición en la misma red hogareña privada a las 14hs. En ésta instancia el trafico de la red fue mucho mayor, ya que 7 dispositivos hicieron uso de la red: 3 computadoras, 3 celulares y un televisor se encontraban conectados a internet via wifi. En ésta ocasión, el alfabeto expuesto fue el siguiente:

$$A = \{(U, ARP), (B, ARP), (U, IP), (B, IP), (U, 35130), (U, IPv6)\}$$

En ésta ocasión, como expone la figura 4, la proporción de paquetes con protocolo IPv6 casi se triplico, la de los IPv4 se redujo por casi la mitad y no se capturaron paquetes PNAC. La proporción de paquetes 35130 y ARP se volvio casi despreciable.

En la figura 5 se puede observar que practicamente todos los paquetes tuvieron destino UNICAST. Solo el 0.664 % (166) tuvieron destino BROADCAST y entre estos 30 tuvieron protocolo ARP mientras que 136 IPv6. Entre los 54 paquetes ARP, 24 tuvieron destino UNICAST, haciendo que el simbolo (U, ARP) fuera el que mas información proveyera.

En el último grafico 6 se puede ver que los paquetes ARP son los que mas información proveen, seguidos por el simbolo (B,IP) y (U,35130).

La entropía tuvo un valor 0.78911, exponiendo que la incertidumbre fue menor que en la medición expuesta anteriormente, y muy lejana al maximo teórico 2.807.

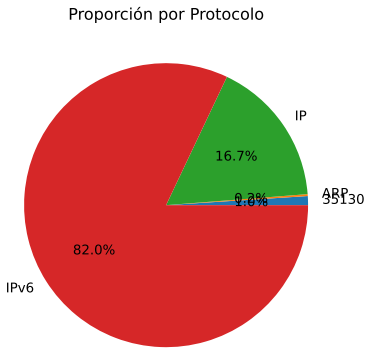

Figura 4: Proporción de paquetes por protocolo.

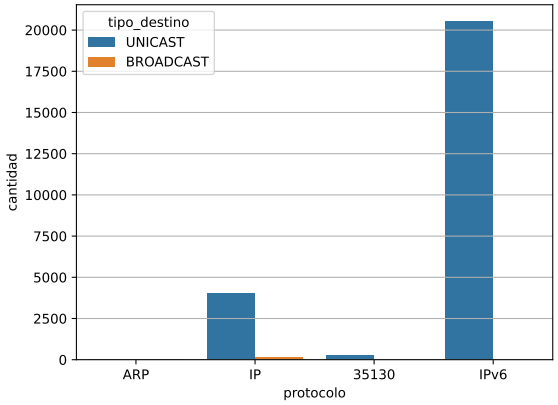

Figura 5: Proporción de paquetes de broadcast y unicast con respecto a los protocolos.

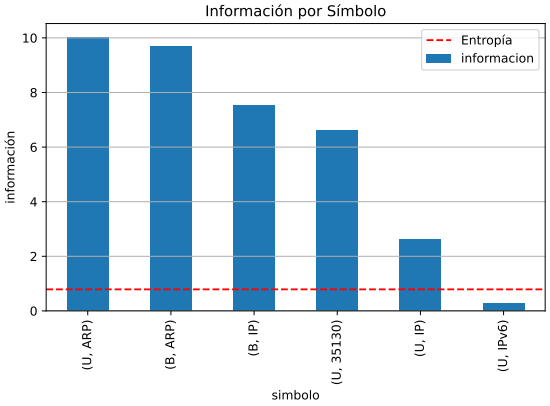

Figura 6: Cantidad de informacion por simbolo comparado con el valor de la entropia.

3.3. Red UBA-WIFI Pabellón 1 - 17hs

La siguientes dos mediciones pertenecen a la misma red estudiada, que fue la de UBA-WIFI en el pabellón 1 de la facultad, aunque con algunas horas de diferencia entre cada análisis, ya que nos interesaba medir como iría a variar el tráfico a lo largo del tiempo. En cada caso se capturaron 12.500 paquetes. Siguiendo la línea de considerar que símbolos conforman nuestra fuente, en este primer caso obtuvimos:

$$A = \{(U, IP), (B, IP), (U, IPv6), (U, ARP)\}$$

con U := UNICAST y B := BROADCAST.
En la figura 7 se muestra la cantidad de paquetes IP que se usan por protocolo. La tasa de paquetes IP es casi el 100 %. Evidentemente, los otros protocolos que aparecieron aportaron mucha más información dado a su rareza en este caso.
Luego, en el gráfico 8 lo que vemos es algo nuevamente ligado a los protocolos, pero ahora nos quedamos con cuantos de ellos son BROADCAST y cuantos UNICAST, donde se visualiza que solo el protocolo IPv4 presento paquetes broadcasteados. Los de tipo BROADCAST pasaron a

representar el 11.464 % del total (1433), casi el doble de lo que representaban en la red anterior. Predominan los UNICAST, con un 88.536 % (11067) de los 12500 paquetes totales.

Por último, se observa en el gráfico 9 lo que se podía imaginar, y es la masiva información que da el último simbolo y casi nada que da el primero. La incertidumbre fue baja y por eso hubo solo 0.65343 de entropía.

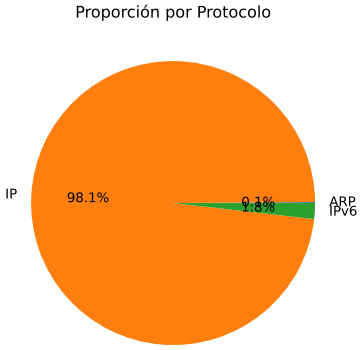

Figura 7: Proporción de paquetes por protocolo.

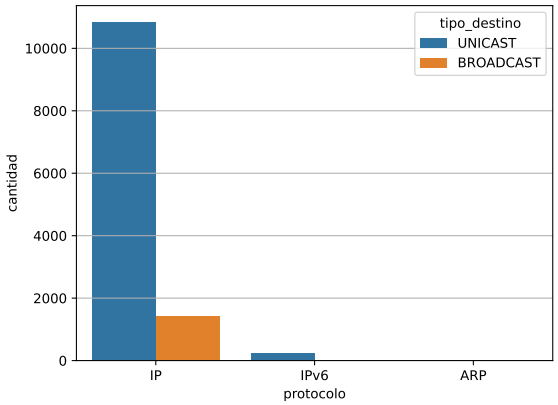

Figura 8: Proporción de paquetes de broadcast y unicast con respecto a los protocolos.

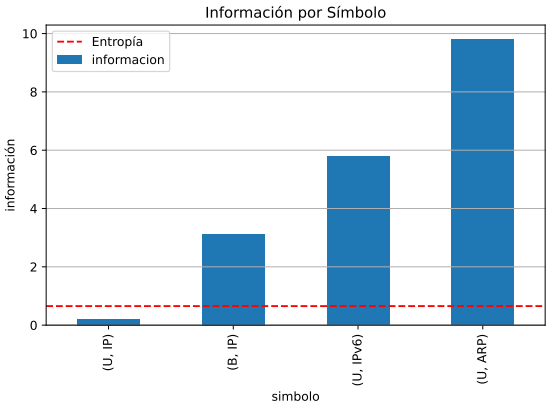

Figura 9: Cantidad de información por símbolo comparado con el valor de la entropia.

3.4. Red UBA-WIFI Pabellón 1 - 20hs

Horas después de la medición anterior, realizamos esta otra, de también 12500 paquetes, en la misma red. Lo que obtuvimos fue más interesante y respondió a esa variación que queríamos observar en una red pública.

Esto se vio reflejado en el aumento de los paquetes de Broadcast, entendiendo también la aparición de un nuevo simbolo que se suma a la fuente respecto a las mediciones anteriores, que es el de (BROADCAST,ARP).

Así todo, observamos en la figura 10 que la tasa de paquetes IP en relación al total se mantuvo como predominante.

Como adelantábamos, lo más interesante se halló en como cambió respecto a las medidas anterio-

res el total de paquetes BROADCAST. Es en el gráfico 11 que observamos este fenómeno.

El total de paquetes UNICAST bajó al 71.408 % (8926), casi un 20 % menos que 3 horas antes, llegando los BROADCAST al 28.592 %, con 3574.
Esta reducción en los paquetes (UNICAST,IP) que prácticamente homogeneizaron la anterior medición, junto al aumento de paquetes BROADCAST con el mismo protocolo, hicieron que las probabilidades de los simbolos sean más diversas, aumentando la cantidad promedio de información. Por eso es que la entropía casi se duplicó para llegar a un 1.219, aunque muy lejos de la máxima teórica de 2.321. La relación de la información con respecto a la entropía se puede ver en la figura 12.

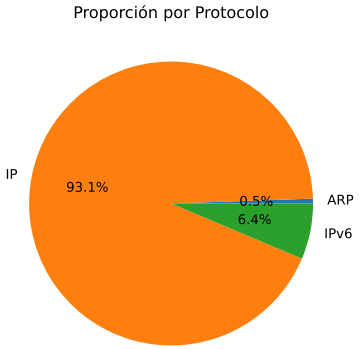

Figura 10: Proporción de paquetes por protocolo.

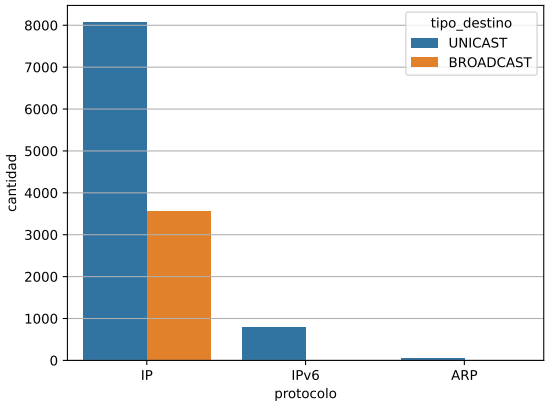

Figura 11: Proporción de paquetes de broadcast y unicast con respecto a los protocolos.

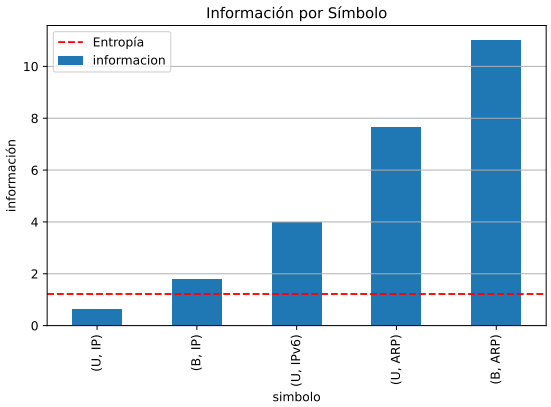

Figura 12: Cantidad de informacion por simbolo comparado con el valor de la entropia.

3.5. Nodos distinguidos

La red presentó un alfabeto compuesto por los siguientes símbolos:

$$A = \{10.100.96.1, 10.100.105.42, 10.100.107.97, 170.51.247.9, 20.201.28.151\}$$

En la figura 13 se observa que el dispositivo distinguido fue el de dirección IP 10.100.96.1, ya que 48 % de los paquetes capturados estuvieron direccionados al mismo. El siguiente nodo distinguido fue el 10.100.105.42, con un 35 % de los paquetes, perteneciente a la computadora portatil que en el momento tomó las mediciones.

La entropía de la fuente fue 1.61, lejano del máximo valor teórico 2.32.

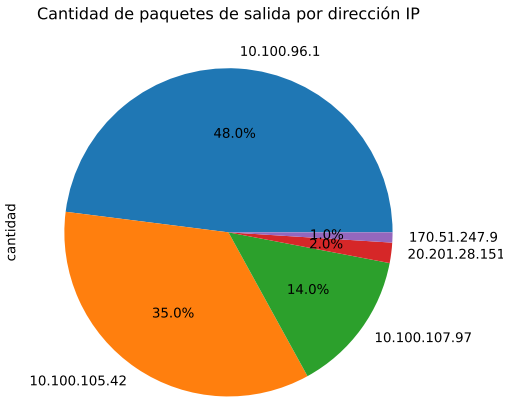

Figura 13: Proporción de paquetes direccionados a cada direccion IP.

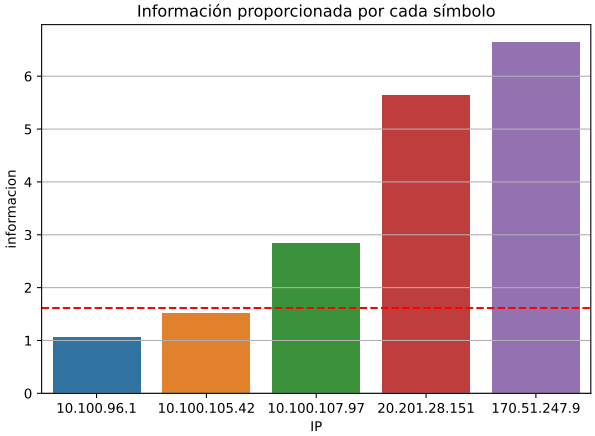

Figura 14: Información de cada direccion IP en comparación con la entropía total.

4. Conclusiones

Como comentario general de los cuatro contextos analizados, se puede observar mayor cantidad de paquetes Unicast en comparación a Broadcast, siendo la mayor parte de estos los involucrados en los protocolos IP. Esto es razonable al considerar que los experimentos se basan en escuchar redes conectadas a Internet.

Para el primer contexto de uso, la medición se lleva a cabo a tal horario que casi ningún host conectado a la red la utilizo activamente, minimizando el trafico en la red y dando a observar sólo los paquetes necesarios para mantener la conexión, y aquellos involucrados en los procesos en segundo plano. Luego, se tiene, en menor medida, los símbolos relacionados al protocolo ARP y PNCAC, pero nos sorprendió llamativamente la proporción del protocolo 35130. Es interesante observar que (ARP, Broadcast) aporta menos información que en los otros casos.

En el horario cotidiano de la red hogareña, es significativa la diferencia entre la cantidad de paquetes IPv6 y los paquetes IPv4. En particular, muestra que el símbolo que menos información brinda es precisamente (IPv6, Unicast), que es contraintuitivo debido a que este protocolo es más moderno y, por ende menos utilizado. Esto podría suceder, sí, por alguna razón, se hubiera accedido a mayor cantidad de servidores que utilizan el protocolo IPv6.

En segundo lugar, las mediciones en la red de la UBA tanto en horario pico como en horario menos concurrido proporcionaron resultados similares: hay mayor cantidad de símbolos relacionados a los protocolos IP, en particular de unicast. Esto se debe a que los experimentos que se llevaron a cabo implica mandar y recibir muchos paquetes de internet. Notamos que es ínfimo el porcentaje que ocupa el protocolo ARP en estas redes, como consecuencia contienen mucha información haciendo que aumente el valor de la entropía.

En relación a la entropía, llegamos a la conclusión que mientras más datos se midieron la entropía se fue estabilizando ya que tener mayor muestra permite estimar mejor el valor real de la entropía de las fuentes.

Podemos ver que la entropía encontrada está muy por debajo en comparación al caso en que el contexto fuera equiprobable, ya que el valor maximo teórico es $H_{\text{máx}} = \log_2(5) \approx 2{,}3219$.

Notando que las entropías observadas están muy por debajo. Esto se debe a que, como se analizó antes, se vió predominancia de ciertos símbolos que aportan poca información.

Con respecto a la experimentación del modelo de nodos distinguidos, primero se puede notar que la entropía observada difiere mucho de la máxima teórica 2.3219, sugiriendo que las direcciones de los paquetes no varían demasiado, haciendo que los paquetes esten direccionados siempre a un minimo subconjunto de direcciones. La única dirección IP cuyo dispositivo era conocido fue la del segundo nodo mas distinguido, sin embargo lo esperado es que el nodo distinguido 10.100.96.1 fuera la dirección IP del mismo router de la red, debido a que es el único dispositivo que se lo direcciona al responder paquetes (Broadcast, ARP).

El resto de las direcciones IP concluimos que pertenecen a hosts, donde la proporción de paquetes dirigidos a los mismos es proporcional al uso que le estaban dando a la red. Sin embargo nos sorprendió que se presentaron pocos simbolos: en una red publica de una facultad se esperó mas trafico en la red y como consecuencia mas simbolos. De todas formas, existe la posibilidad de que el router haya tenido la mayoría de las direcciones IP ya mapeadas en la memoria cache y como consecuenci reduciendo considerablemente el trafico de paquetes ARP.

Referencias

[1] ÏEEE 802 Numbers,Ïnternet Assigned Numbers Authority (IANA). [Online]. Available: https://www.iana.org/assignments/ieee-802-numbers/ieee-802-numbers.xhtml. [Accessed: 13- Sep- 2023].