Protocolo ARP

December 6, 2014

Universidad de Buenos Aires - Departamento de Computación - FCEN

Integrantes:

- Gallardo, Guillermo L.U.: 32/10 gagdiez.c@gmail.com
- Fixman, Martin L.U.: 391/11 martinfixman@gmail.com
- Matayoshi, Leandro L.U.: 79/11 leandro.matayoshi@gmail.com
- Szyrej, Alexander L.U.: 642/11 alexander.szyrej@gmail.com

Contents

1 Introducción

El protocolo ARP (*Adress Resolution Protocol*) permite mapear direcciones IP a direcciones físicas. Cuando un host desea comunicarse con otro dentro de una misma red local debe hacerlo utilizando la *MAC Adress* del mismo. Conociendo su dirección IP, puede *brodcastear* un paquete *ARP who-has* pidiendo la *MAC* asociada a dicha IP. De esta manera recibirá en respuesta un paquete *ARP-reply* por parte del host destino. Para evitar una cantidad excesiva de broadcasts, cada nodo de la red mantiene su propia tabla de mapeos que actualiza periódicamente.

Cada paquete ARP incluye numerosos campos. Sin embargo, pondremos énfasis especialmente en los siguientes: Operation (indica si es un pedido o respuesta), Sender Hardware Adress, Sender Protocol Adress, Target Hardware Adress y Target Protocol Adress.

El objetivo de este trabajo práctico es analizar el flujo de paquetes *who-has* en distintas redes públicas y privadas. Luego, utilizando como base el concepto de *entropía de la información* analizaremos el rol que juega cada dispositivo dentro de la red.

2 Métodos

Sea F una fuente de información de simbolos finitos S, esto es, F produce periódicamente algún $s_i \in S$. Supongamos conocemos las probabilidades $P(s)$ con que se produce cada símbolo $s \in S$. Se define la *Cantidad de información* de un símbolo s como:

$$\frac{I(s) = log_2(1}{P(s))}$$

Se define la *entropía de la fuente* como:

$$H(S) = \sum_{s \in S} P(s)I(s) = -\sum_{s \in S} P(s)log_2(P(s))$$

La *Entropía* se considera una medida de incertidumbre, esto es, a mayor valor, menor homegeneidad habrá en los mensajes emitidos por la fuente. Es importante notar que $0 <= H(S) <= log_2(\#S)$. La *Entropía* será nula cuando la fuente emita siempre el mismo símbolo entre todos los posibles, mientras que se maximiza cuando todos los simbolos son emitidos equiprobablemente.

Dado que los símbolos definen a una fuente, uno puede modelar una misma red como distintas fuentes. Luego, utilizando el teorema central del límite es posible estimar la probabilidad de cada símbolo utilizando su frecuencia de aparición. Esto permite calcular la entropía en redes informáticas.

3 Desarrollo

El primer paso en este trabajo fue implementar un script en *python* para escuchar pasivamente paquetes dentro de una red local y capturar información sobre el tráfico de la red.

Esto lo hacemos capturando todos los paquetes que circulan utilizando la librería Scapy, y nos quedamos solo con los ARP de tipo *who-has*. Luego, contamos cuantas veces aparece una dirección en cada uno de los campos *dst* y *src*, así como también cuantos paquetes envío cada IP al resto.

Con estas observaciones calculamos la entropía de la red modelandola como dos fuentes distintas: en la primera, denominada S_{src}, los símbolos son las direcciones IP que aparecen en el campo SRC de los paquetes capturados; en S_{dst} son las direcciones IP que aparecen en el campo DST.

El script se utilizó para analizar cuatro redes, dos privadas controladas, y dos públicas no controladas. Decimos que una red es controlada si durante la captura de paquetes sabemos a qué nodo corresponde cada ip. Las redes privadas son una casa de familia y una red empresarial, las públicas son la red *Entrepiso DC* de la facultad y un local de Starbucks.

4 Resultados y Análisis

4.1 Casa Familia

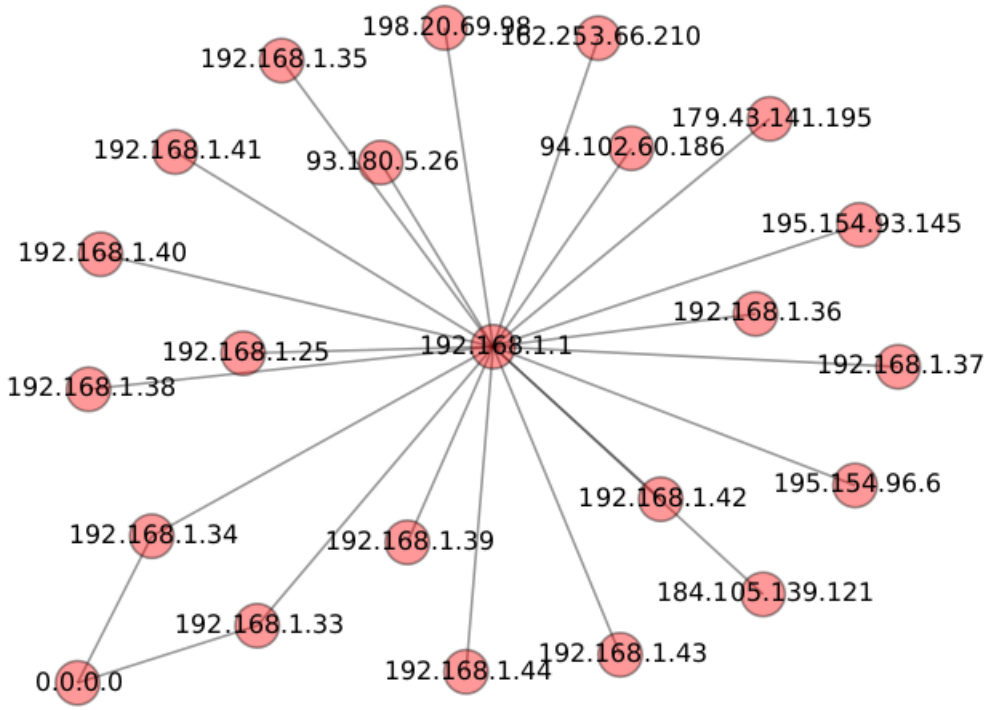
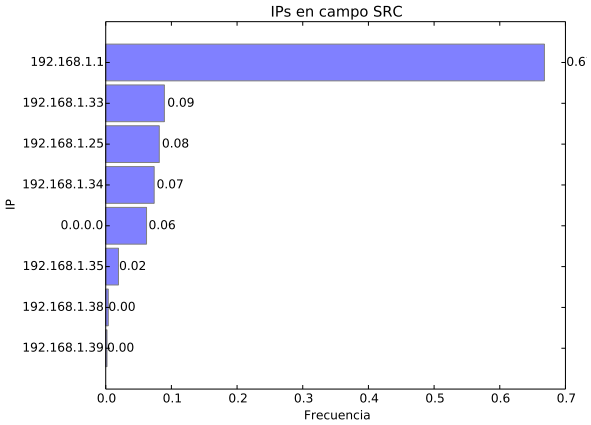


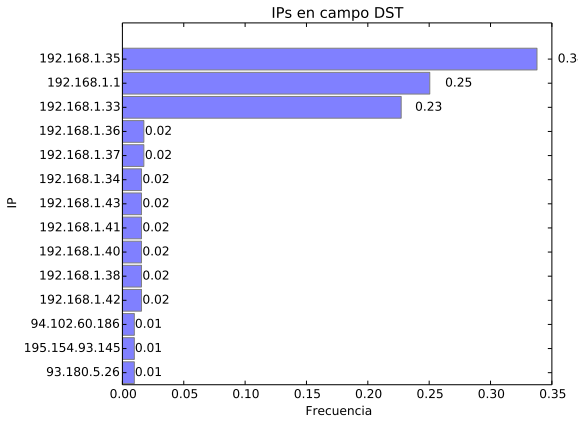
Figure 1: Grafo de conectividad en la casa

En el caso de la casa familiar sabemos que la dirección del router es 192.168.1.1 y que 192.168.1.35 pertenece al dispositivo que más tiempo estuvo prendido, el mismo es una notebook. Podemos apreciar que la ip del router es de las que aparece con mayor frecuencia en los campos SRC (fig *a*) o DST (fig *b*), mientras que la notebook solo aparece un alto número de veces en DST.

En la *Figura 1* representamos la red como un grafo, en el mismo los nodos son dispositivos y las aristas representan que existió al menos un paquete que tenía a alguno de ellos como *src* y al otro como *dst*. Podemos ver que la red quedó representada con el router en el centro. La *Figura 2* presenta esta misma información pero con mayor detalle, nuevamente los nodos son dispositivos, pero en este caso las aristas son dirigidas, cada arista representa la cantidad de veces que el nodo origen envió un paquete ARP al nodo destino.



a. Estimación de la probabilidad de cada símbolo en modelo SRC



b. Estimación de la probabilidad de cada símbolo en modelo DST

La entropía fue: 1.68 para el modelo SRC y 2.67 para el modelo DST. Teniendo en cuenta que se vieron 23 ips asumimos que existen 23 símbolos posibles para cada fuente, con lo cual podemos calcular el límite superior para la entropía de estos modelos, la misma es $log_2(23) = 4.52$, lo cual hace que las entropías normalizadas sean aprox. 0.37 para SRC y 0.59 para DST. Esto indica que es mas probable encontrar la mismas IP enviando siempre paquetes, mientras que las IP que reciben la mayor cantidad de paquetes son varias. Lo que ya se veía reflejado en los gráficos *a y b* donde el router es el único que posee una frecuencia de envío 6 veces superior a todo el resto de los nodos, por otro lado en DST son tres los nodos en esta situación.

Cabe destacar algo interesante que sucedió, aparecieron paquetes ARP que poseían direcciones que no pertenecian a la red local.

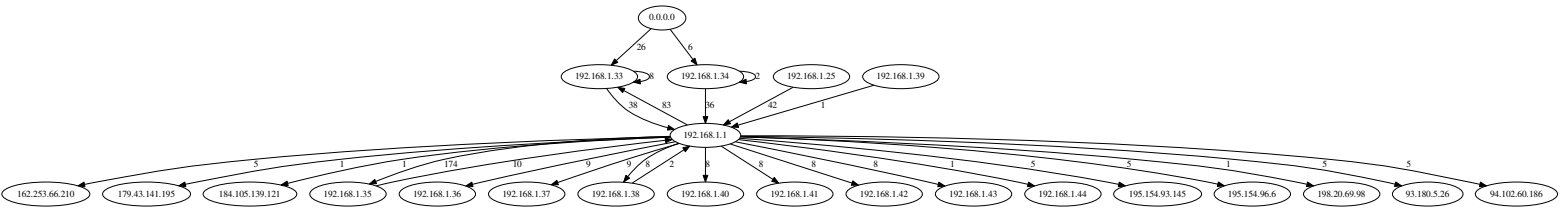


Figure 3: Intercambio de paquetes ARP en la casa

4.2 Empresa

Al sniffear la empresa se vieron en total 40 ips distintas, para poder visualizar los datos nos quedamos solo con los 23 pares de nodos que mayor intercambio de paquetes ARP tuvieron entre ellos. Nuevamente representamos la red como grafos (*Figuras 4 y 5*), siguiendo los mismos procedimientos que en el caso anterior. Podemos ver que nuevamente el router aparece en el centro de los grafos, aunque es posible observar una interacción mucho mayor entre distintos pares hosts, y no únicamente entre cada uno de estos y el router.

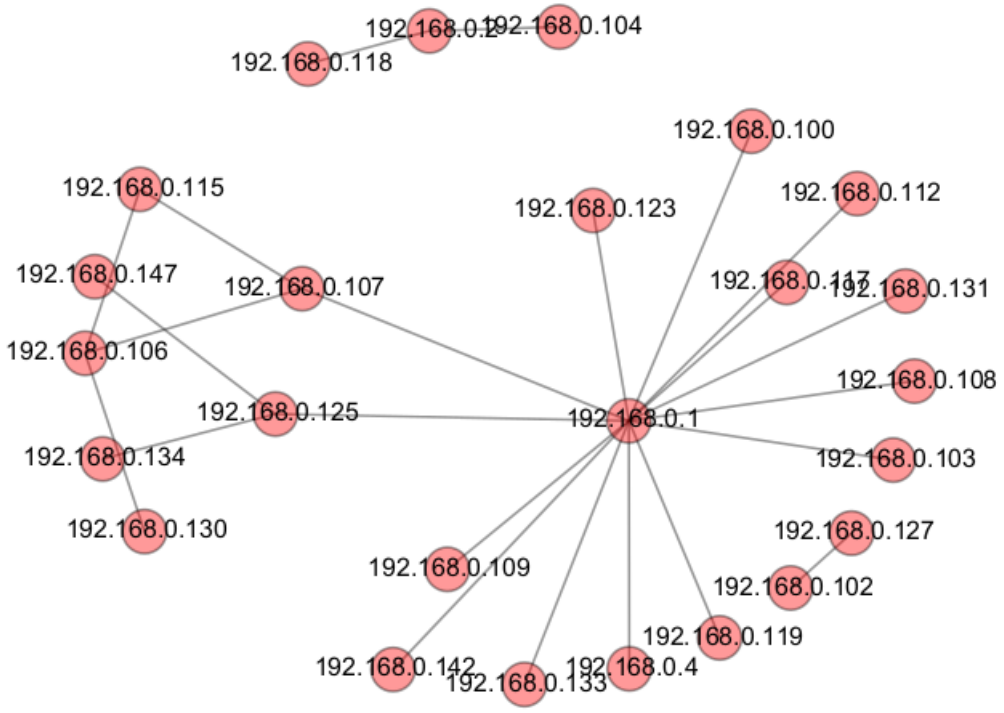
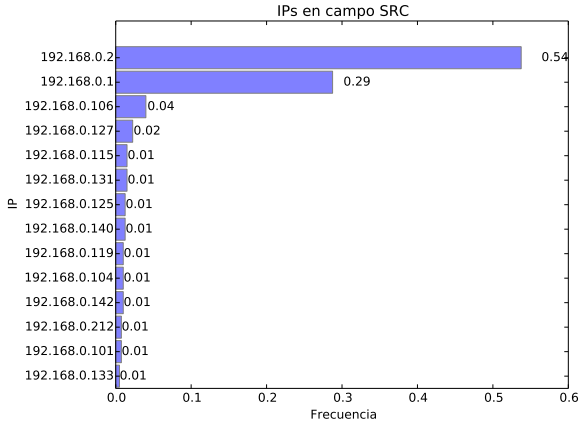
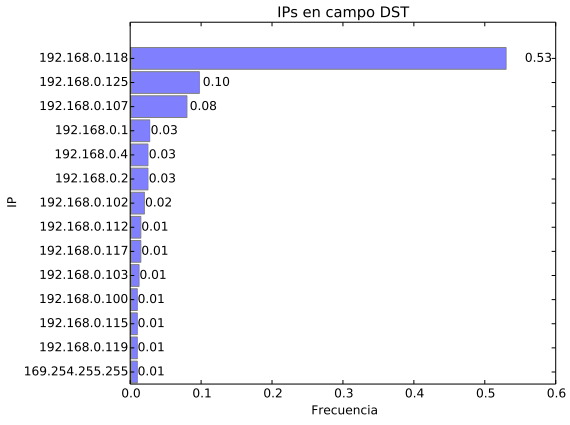


Figure 4: Grafo de conectividad en la empresa



c. Estimación de la probabilidad de cada símbolo en modelo SRC



d. Estimación de la probabilidad de cada símbolo en modelo DST

El valor de la entropía para el modelo SRC es 2.05. El valor de la entropía para el modelo DST es 2.99. Teniendo en cuenta que se vieron 40 ips, asumimos que las fuentes poseen 40 símbolos, lo cual nos permite calcular la entropía maxima ($log_2(40) = 5.32$) lo cual normaliza las entropias a 0.38 y 0.56. Nuevamente el modelo SRC posee un valor mayor.

En el gráfico de barras del modelo SRC (gráfico c) el router juega un papel relevante ya que es el segundo símbolo con mayor cantidad de emisión de paquetes. Es interesante mencionar que el nodo con mayor cantidad de paquetes enviados, cuya ip es 192.168.0.2, corresponde a un servidor apache. El resto de los nodos que aparecen en el gráfico corresponden a notebooks y celulares conectados a la red, ya a partir del quinto elemento aparecen con una frecuencia de aparición hasta 50 veces mas baja.

En el gráfico de barras del modelo DST (gráfico d) es en donde se nota la mayor cantidad de diferencias.

El router no se corresponde con ninguno de los 3 nodos con mayor cantidad de paquetes recibidos:

- 192.168.0.118 corresponde a una notebook
- 192.168.0.125 corresponde a un servidor apache
- 192.168.0.107 corresponde a un servidor apache

Recién aparece en la cuarta posición, con una cantidad de paquetes recibidos considerablemente menor a los primeros. En esta red en particular los dispositivos se envían paquetes who-has entre ellos en mayor medida, y hacia los servidores en vez de hacia el router. A su vez, podemos ver que en este caso son ocho las ips que poseen desde el doble hasta 50 veces mas frecuencia de aparición. Suponemos que por esto la entropía de este modelo es mas alta.

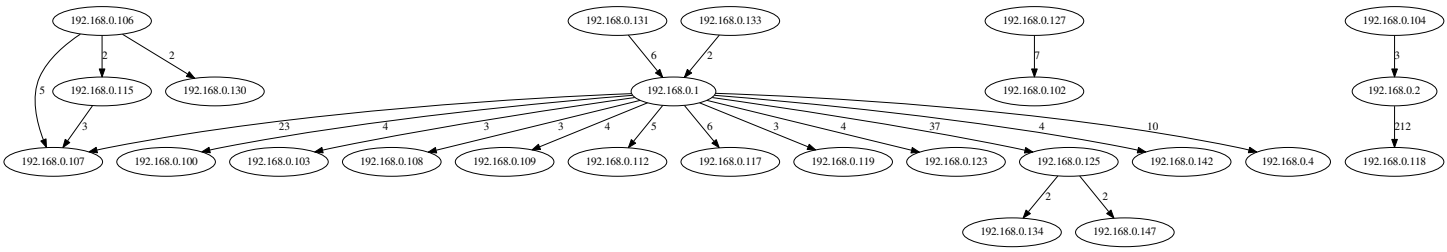


Figure 6: Intercambio de paquetes ARP en la empresa

4.3 Starbucks

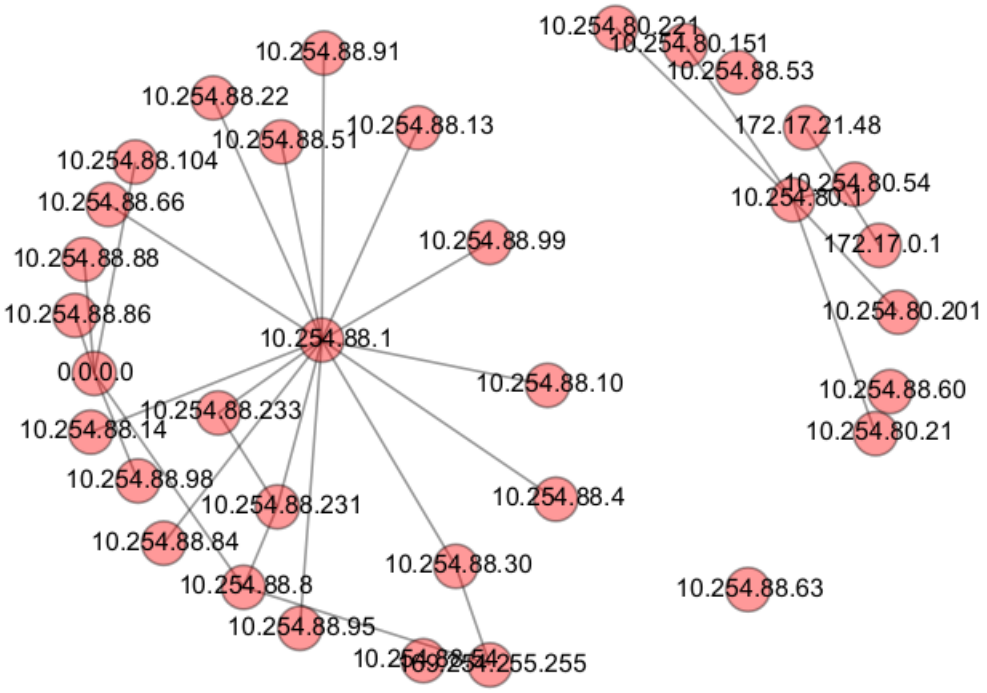
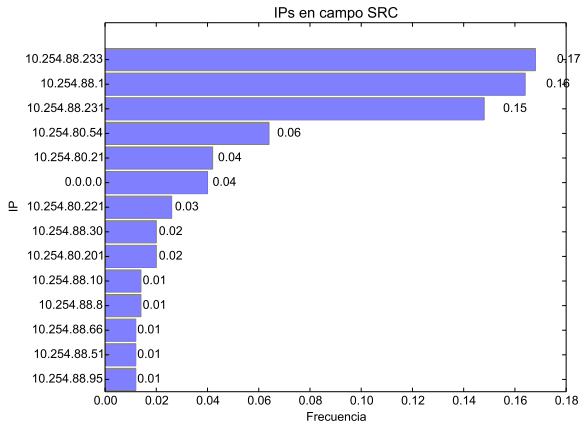
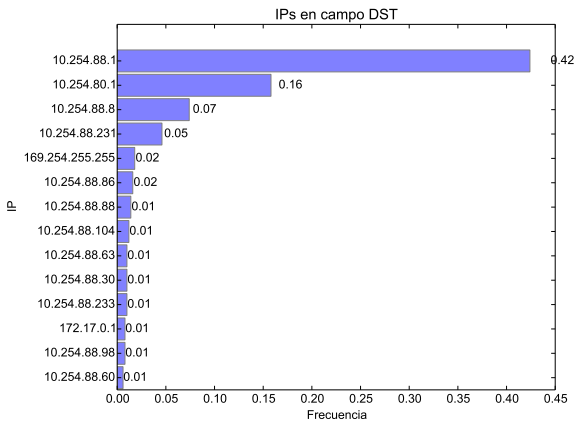


Figure 7: Grafo de conectividad en Starbucks

Para el caso de la red abierta disponible en Starbucks se recopiló información de 200 nodos en total, la *Figura 7* fue creada con solo 40 de ellos, en caso de agregar más nodos se notaría que todos están conectados que el nodo 10.254.88.1. A su vez posee una alta frecuencia de aparición en ambos modelos. Esto sería el comportamiento esperado del router de la red 10.254.88.0/24. Asumimos que ésta es la dirección de la red puesto que la mayoría de las IPs de los paquetes capturados difieren en el último octeto. Por otro lado, podemos ver que la segunda dirección con mayor frecuencia en DST fue 10.254.80.1, seguida por 10.254.88.8. Suponemos que 88.8 es la pc del lugar dada la ip baja, y la cantidad de apariciones. Por otro lado, no sabemos que es 80.1 dado que parece formar una red aparte de la sniffeada.



e. Estimación de la probabilidad de cada símbolo en modelo SRC



f. Estimación de la probabilidad de cada símbolo en modelo DST

La entropía fue: 4.61 para el modelo *SRC* y 3.76 para el modelo DST. La entropía máxima dado los 200 símbolos es $log_2(200) = 7.64$, por lo que normalizadas quedarían $SRC = 0.60$ y $DST = 0.48$. Esto es consistente con los graficos (e y f) donde las frecuencias de aparición de cada símbolo decaen mucho más rápidamente en DST.

Nuevamente entre los paquetes capturados volvieron a aparecer direcciones que no parecieran pertenecer a la red local, resultando sumamente interesante el caso de 10.254.80.1

4.4 Entrepiso

En el caso del entrepiso se detectaron 205 ips, lamentablemente la forma en que las mismas se comunicaron entre si hizo imposible el graficar una representación adecuada con las herramientas que usamos. La *Figura 9* muestra varios subgrafos conexos, donde se destacan los nodos: 10.1.100.254, 10.1.200.30 y 10.1.200.254. Por las direcciones IP 200.254 y 100.254 parecieran ser routers. La entropía fue: 4.25 para el modelo *SRC* y 4.90 para el modelo DST. Estos valores normalizados son: 0.55 y 0.63 respectivamente, ambos son valores muy altos. Podemos ver esto reflejado en los gráficos *g y h*, donde las diferencias en frecuencias entre los nodos no es tan grande como en los casos vistos hasta ahora.

A su vez, si no fuera por la *Figura 8* no podriamos buscar un candidato tan claro a router, ya que 200.254 y 100.254 aparecen recién en cuarto y quinto lugar para SRC, mientras que aparecen en quinto y decimo lugar para DST.

Una vez mas aparecieron direcciones sueltas, como es por ej: 10.1.11.254.

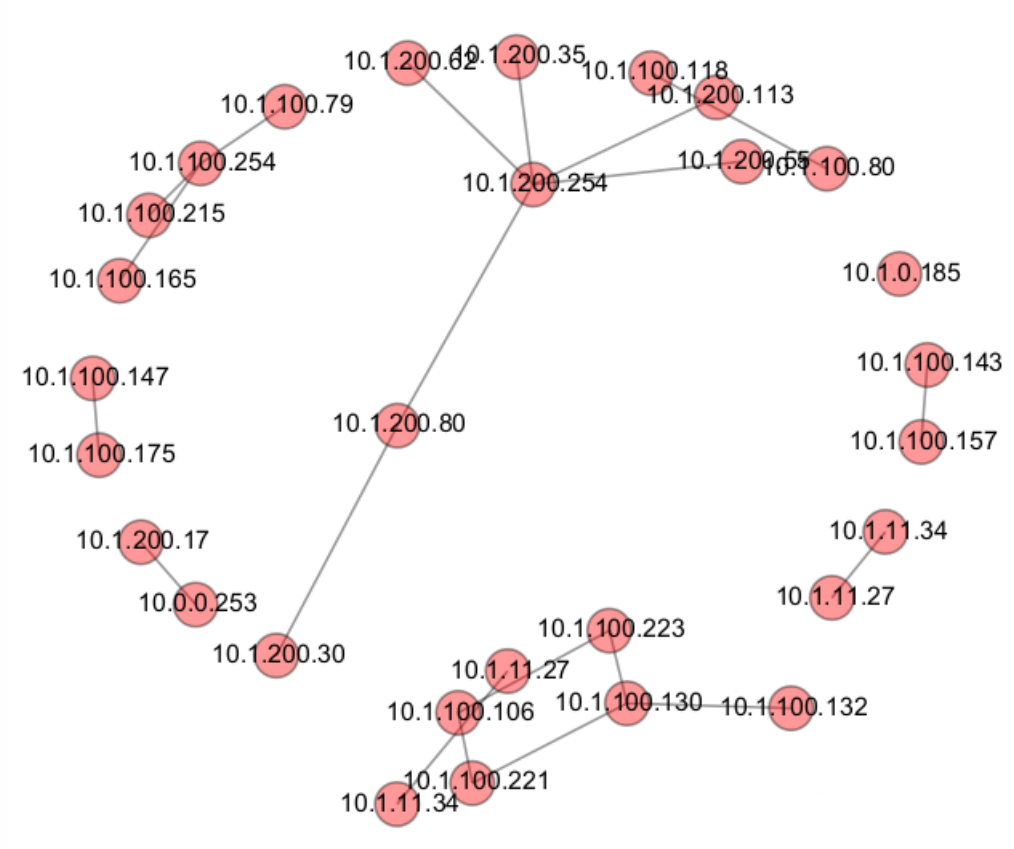
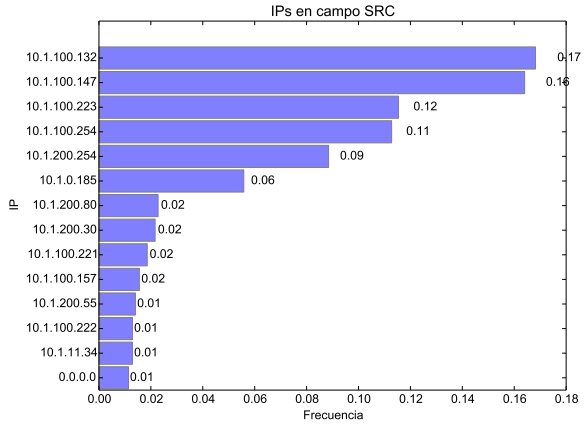
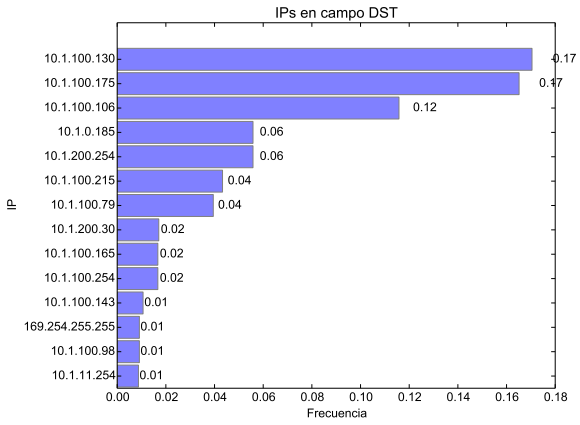


Figure 9: Grafo de conectividad en la red Entrepiso



g. Estimación de la probabilidad de cada símbolo en modelo SRC



h. Estimación de la probabilidad de cada símbolo en modelo DST

4.5 Discusión

En todas las redes sucedió que aparecieran direcciones ip que se supone no pertenecen a la misma.

En algunos casos apareció la dirección 169.254.255.255. Investigando encontramos que es utilizada como broadcast por DHCP que es un protocolo de configuración automática de parámetros de red tales como direcciones IP para interfaces y servicios.

La dirección 0.0.0.0 es el estándar para *broadcastear* dentro de una red local.

Por otra parte, el resto de las direcciones IP que aparecen pueden deberse a que los routers posean la opción de *Proxy ARP* habilitada, cosa que tiene mucho sentido en el caso del Entrepiso por ejemplo.

Es posible que las direcciones que aparecen con mayor frecuencia en el grafo de pedidos ARP correspondan los dispositivos que estabam más activos durante el periodo de tiempo que se hizo el sniffing.

5 Conclusiones

Capturando paquetes who-has de una red es posible detectar los nodos relevantes de una red, por ej, en todos los casos los routers resultaron ser facilmente detectables al representar en forma de grafo los intercambios de paquetes ARP por parte de pares de nodos, dado que los mismos son quienes interactuan con la mayor cantidad de nodos.

Por otra parte, el analizar la frecuencia de aparición de las ips en cada campo junto con el cálculo de la entropía en la red permitió demostrar empíricamente que la entropía es mas baja cuando son pocos los símbolos que aparecen con mucha frecuencia. A su vez, vimos que de acuerdo a como modelaramos la fuente los resultados eran distintos. Pero no solo los resultados en valores, en los modelos SRC el router siempre ocupó una posición menor que en el modelo DST. Esto puede ser de ayuda para localizar dispositivos importantes en redes de entropía baja, por ej, en la casa y la empresa. En redes mas complejas como la de starbucks, no aportó ninguna información mas que demostrar que la red era caótica.

6 Referencias

- **Computer Networks: A Systems Approach**, *Larry L. Peterson and Bruce S. Davie.*
- **Computer Networks**, *Andrew S. Tanenbaum*
- **Special-Purpose IP Address Registries** (RFC 6890), *Internet Engineering Task Force (IETF)*
- **An Ethernet Address Resolution Protocol or Converting Network Protocol Addresses to 48.bit Ethernet Address for Transmission on Ethernet Hardware** (RFC 826), *David C. Plummer*