DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Sanguijuelas

2 de septiembre de 2014

Métodos Numéricos
Trabajo Práctico Nro. 1

Integrante	LU	Correo electrónico
Martin Carreiro	45/10	martin301290@gmail.com
Kevin Kujawski	459/10	kevinkuja@gmail.com
Juan Manuel Ortíz de Zárate	403/10	jmanuoz@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) 4576-3359
http://www.fcen.uba.ar

Índice

1. Resumen

Mediante el manejo de matrices se buscará modelar y solucionar una problema de la realidad.Cómo esta esta compuesta de infinitas variables dicha modelización implicará una inevitable discretización. Es decir trabajar con una cantidad acotadas de variables del problema (sólo las relevantes). Si bien el problema en cuestión consiste en decidir cual/es de las sanguijuelas que están adheridas al parabrisa se debe eliminar la modelización del mismo no es trivial. Es más, podría decirse que este proceso es mucho mas complejo y costoso que la solución en sí. ¿Por qué? porque la creación de la matriz que represente al parabrisa y el cálculo de las temperaturas en cada punto, si no se usa buen método, podría llegar a demorar mucho tiempo, tirar overflows o directamente nunca terminar.

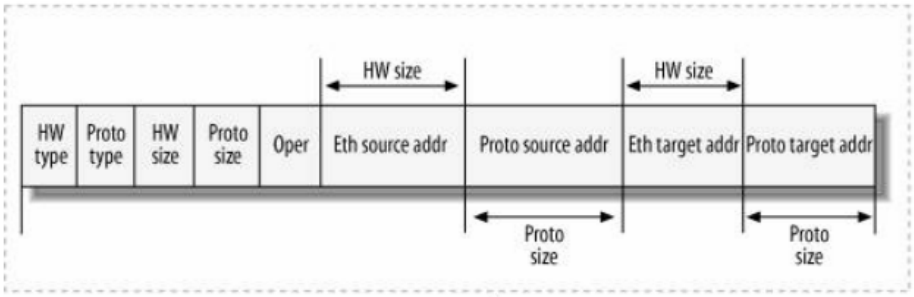
Por esto es que a lo largo de este tp haremos mucho foco en como calcular las temperaturas, como optimizar el espacio ocupado por la matriz obtenida y como optimizar lo mas posible todas las operaciones matriciales.

Palabras clave: ARP, entropía, cantidad de información, fuentes de información

2. Introducción teórica

Para escuchar una determinada red local, se va a utilizar el protocolo ARP (*Address Resolution Protocol*) para capturar los distintos paquetes de este formato enviados a través de la red. ARP es un protocolo que permite mapear direcciones de nivel de red a direcciones físicas. Es decir, establece una relación entre las capa de red y la capa de enlace.

Se necesita un formato de paquete determinado para encapsular este protocolo de resolución de direcciones, para mapear las direcciones de red en direcciones físicas. El paquete posee el siguiente formato:



Los primeros dos campos indican los protocolos de nivel de enlace y nivel de red involucrados en la comunicación. Los siguientes dos campos contienen las longitudes en bytes de cada dirección de hardware y de protocolo. El campo `Oper` puede tomar los valores 1 (*who-has*) o 2 (*reply*): el código *who-has* se refiere a que el emisor del paquete necesita saber la dirección física correspondiente a cierto IP; el código *reply* permite a un dispositivo dar a conocer su dirección de hardware.

Los siguientes campos son las direcciones de red y las direcciones de hardware tanto del dispositivo emisor como del dispositivo destinatario de los paquetes.

Entonces, el objetivo es capturar este tipo de paquetes para analizar la cantidad de información de las distintas IPs de la red que se comunican y calcular la entropía de fuentes de información basadas en los paquetes ARP. Recordar la fórmula para obtener la cantidad de información de un símbolo s:

$I(s) = -log(P(s))$

Y para calcular la entropía de una fuente de información S:

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

3. Desarrollo

Como se mencionó anteriormente, el objetivo es analizar la estructura de algunas redes locales mediante la captura de paquetes ARP que se envían por esas redes. Estos paquetes servirán para obtener información sobre los dispositivos de red y sacar conclusiones de los resultados que se consigan.

Lo primero que se hace en el trabajo es implementar una herramienta para escuchar de manera pasiva una red local. La idea es que la red local que analicemos no sea una red controlada, para favorecer una análisis más rico (si fuera controlada, sabríamos por ejemplo cuál es la IP del router). Para eso se utiliza *Scapy* una biblioteca escrita en `Python`, para la captura y manipulación de paquetes en redes.

El segundo paso es definir los siguientes modelos de fuente de información:

- $S_{src} = \{s_1, ..., s_k\}$ donde cada símbolo s_i es una dirección IP que aparece como dirección origen en los paquetes ARP de tipo *who-has*.
- $S_{dst} = \{d_1, ..., d_k\}$ donde cada símbolo d_i es una dirección IP que aparece como dirección destino en los paquetes ARP de tipo *who-has*.

Con estas fuentes definidas, ya se puede estimar las probabilidades de cada IP que necesitemos de los paquetes ARP capturados y calcular la cantidad de información de cada uno para luego conocer la entropía de cada fuente. Para estimar las probabilidades de las IPs en las fuentes, lo que se hace es simple. Supongamos que `#paquetes` es la cantidad total de paquetes ARP *who-has* obtenidos y `#apariciones` es la cantidad de veces que aparece una *ip* determinada (la explicación sirve para ambas fuentes de información), entonces se define la probabilidad de esa *ip* como:

$$\mathrm{P}(ip) = \frac{\#apariciones}{\#paquetes}$$

Con esta definición se puede calcular la cantidad de información de cada IP obtenida en las fuentes. Luego, es fácil obtener la entropía de cada fuente de información.

Entonces, haciendo uso de esta herramienta implementada, se deben realizar capturas de paquetes ARP sobre distintas LANs (*Local Area Network*) para poder hallar nodos (dispositivos de red) distinguidos. El router que oficia como *gateway* (se comunica con otras redes) de la LAN analizada es de particular interés. La probabilidad de que su IP aparezca en un paquete ARP *who-has*, sea como emisor o como destinatario deberiía ser alta. Es decir, en teoriá debería ser la IP más frecuente en este tipo de paquetes porque es el canal comuń de comunicación. Otra manera de verlo, es la siguiente: se busca la dirección IP (notar que puede haber más de un *gateway*)cuya probabilidades la más alta y por tanto, es la que menor cantidad de información aporta. Se supone que las direcciones IP cuya información sea más cercana al valor de la entropía de la fuente serán los *gateways* de la red.

Asimismo, además de analizar la LAN para poder identificar el o los *gateways*, se monitorea la actividad de las otras IPs que se referencian en la red para registrar la frecuencia de cada una, comparando su información con la entropía de la fuente y quizás, encontrar otros nodos especiales.

4. Resultados

A continuación se presentan los resultados obtenidos de utilizar la herramienta implementada para realizar una escucha pasiva de distintas redes locales y calcular las entropías de las fuentes definidas. Estos experimentos se realizaron en 4 lugares distintos: en el shopping Alto Palermo, en un McDonald's, en un Subway, y por último, en la casa de un miembro del equipo para comparar el tráfico en una red controlada.

4.1. Red del Alto Palermo

Para realizar este experimento, capturamos los paquetes de la red del shopping Alto Palermo. La IP del host que capturaba los paquetes es `172.17.82.139`, y la duración del expermiento fue de 30 minutos. Para ilustrar la comunicación que se estableció en la red durante el experimento, se utiliza un grafo que muestra la interacción entre las IPs con los paquetes que se enviaron. Los nodos del grafo representan las IPs que participaron en la comunicación; los ejes son dirigidos, si existe un eje del nodo IP_1 al nodo IP_2 significa que IP_1 es el emisor de un paquete e IP_2 es el destinatario de dicho paquete. Además, los ejes tienen un peso que es la cantidad de veces que se emitieron paquetes desde una IP a otra.

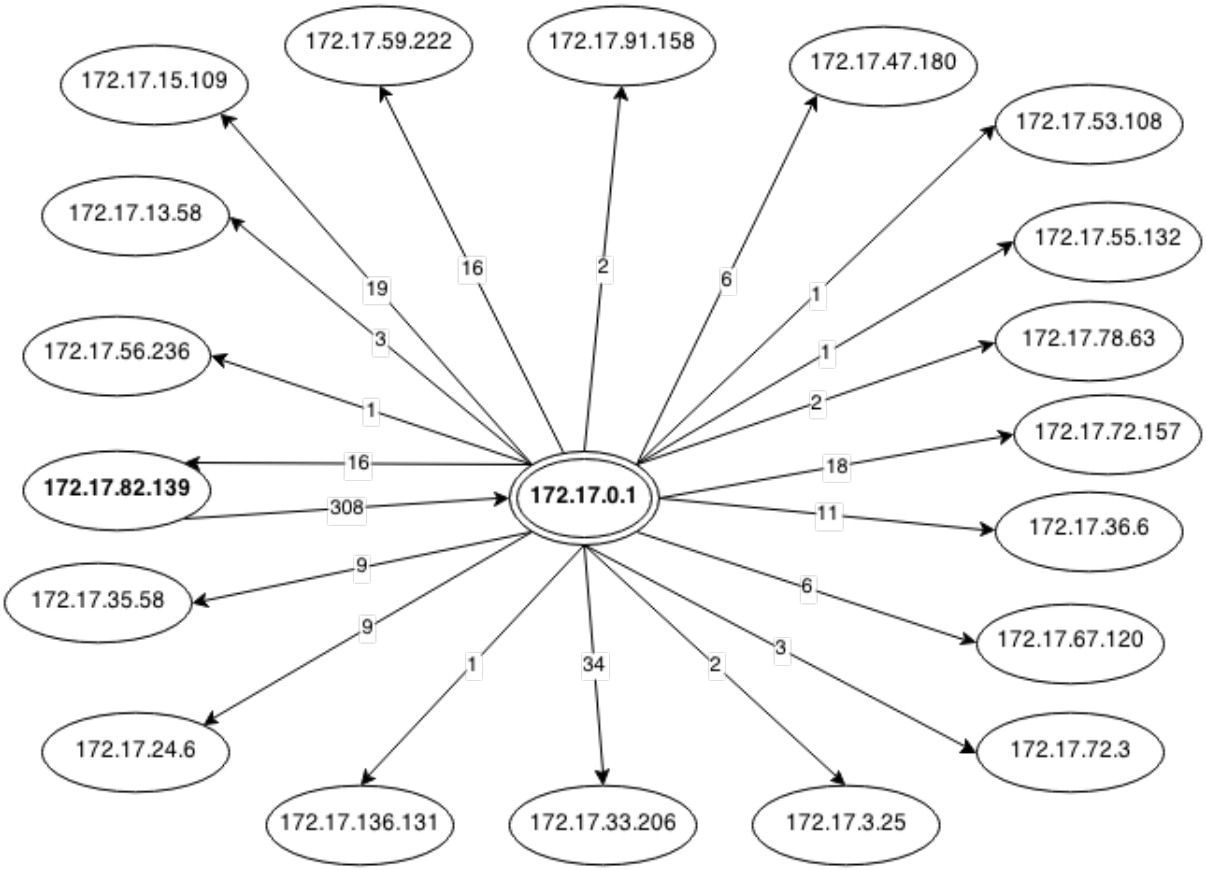


Figura 1: Grafo de paquetes para la red del Alto Palermo

Para comenzar, se supone que la IP `172.17.0.1` corresponde a un router asignado como default *gateway* para la mayoría de los hosts, por su dominio y dado que nuestro host le realiza muchos pedidos y que él se comunica con otros hosts.

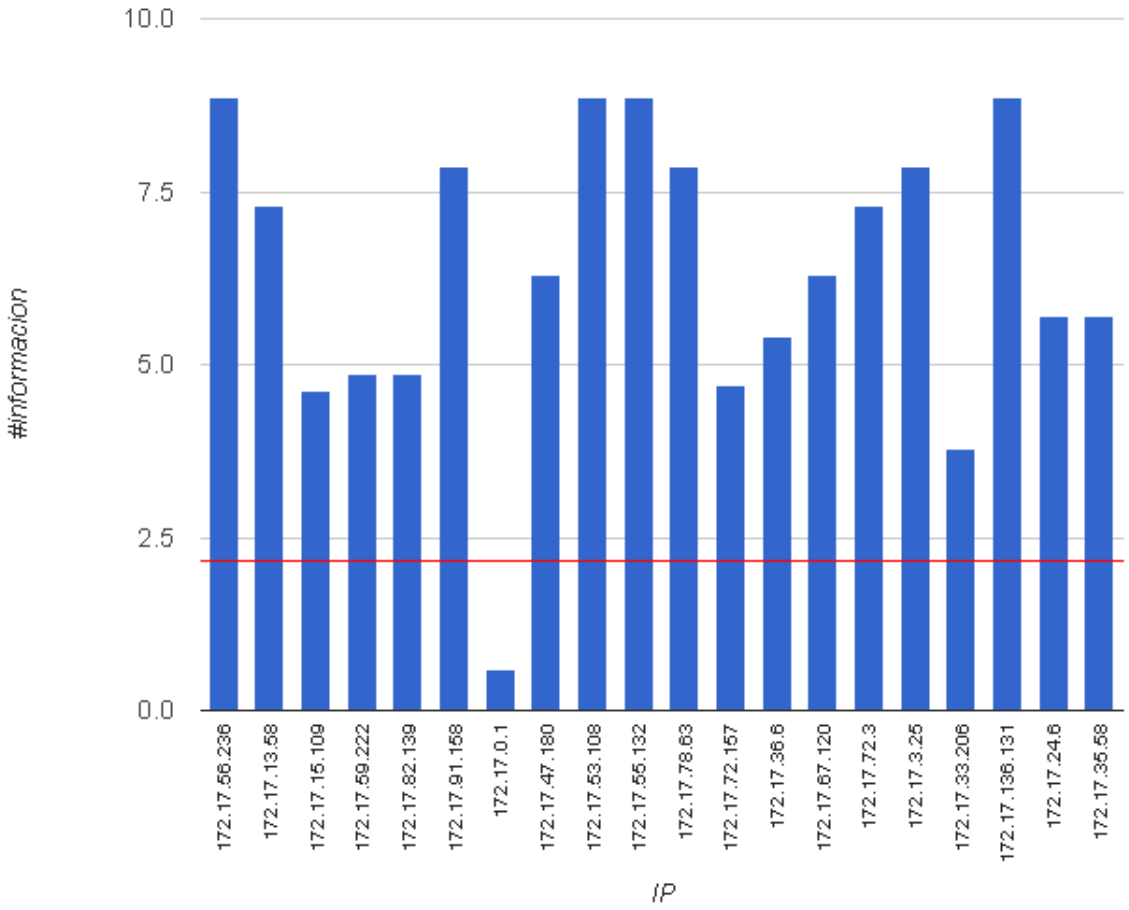


Figura 2: Cantidad de información de cada IP como destinatario, red Alto Palermo

En el gráfico anterior se muestra la cantidad de información por cada IP destino. Es decir, la cantidad de información de cada IP tomando como fuente de información S_{dst} en esta red. Notar que la línea horizontal del gráfico corresponde al valor de la entropía de la fuente. El gráfico sobre la fuente de información S_{src} no aporta muchos datos ya que, a partir del grafo de comunicaciones se puede ver que a los únicos emisores de paquetes son la IP del router y la IP del host. En base a esto y como suponíamos, se aprecia que la menor cantidad de información y la única que se ubica por debajo del valor de entropía (cuyo valor es de 2.15) está dada por la IP asignada al router.

4.2. Red del Subway

En este caso es realizó la captura de paquetes de la red de un Subway. Un aspecto a observar es que la red *wireless* (el Wi-Fi) no funcionaba o no tenía senãl, así que los paquetes ARP capturados están en el contexto de Ethernet. A continuación se muestra el gráfico correspondiente a la cantidad de información de cada IP de la fuente S_{src}, en una captura de una duración de 30 minutos. Nuevamente, la línea roja se refiere al valor de la entropía de la fuente (cuyo valor es 0.4448).

En este caso, el gráfico respecto a la fuente S_dst no tiene mucha relevancia porque el destinatario era la IP `192.168.1.1` con una altísima probabilidad. Es decir, esa IP correspondía al router de la red analizada.

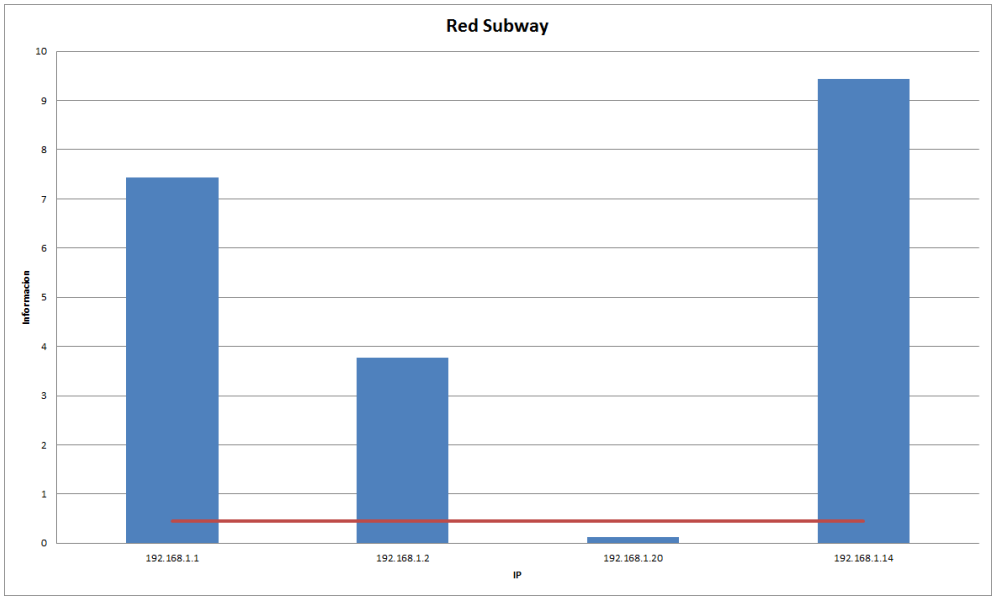


Figura 3: Cantidad de información de cada IP como emisor, red Subway

4.3. Red de Casa

Los paquetes recolectados de esta red representan una captura de 30 minutos aproximadamente de los paquetes ARP *who-has* que fueron enviados por broadcast a través de la red Wi-Fi de la casa de uno de los miembros del equipo (vive en un departamento). EL gráfico corresponde a la cantidad de información de cada IP de la fuente S_{dst}. Recordar que la línea horizontal roja es el valor de la entropía de la fuente.

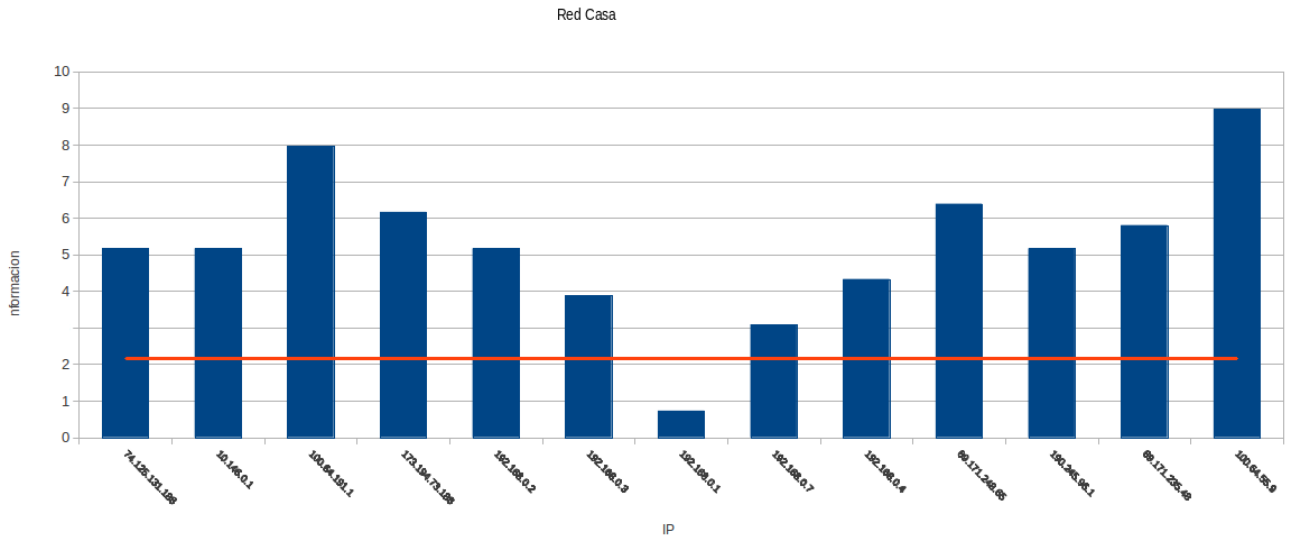


Figura 4: Cantidad de información de cada IP como destinatario, red Casa

4.4. Red del McDonald's

La herramienta implementada también se ejecuto en un McDonald's para observar el tráfico de paquetes que tiene. Este experimento fue el más corto: se dejó corriedo la herramienta 15 minutos. En este experimento

se observó una alta actividad de red, es decir, paquetes enviados en casi todo momento. A continuación se muestra el gráfico de la cantidad de información de cada IP tomando como fuente S_{dst}. Recordar que la línea horizontal roja es el valor de la entropía de la fuente.

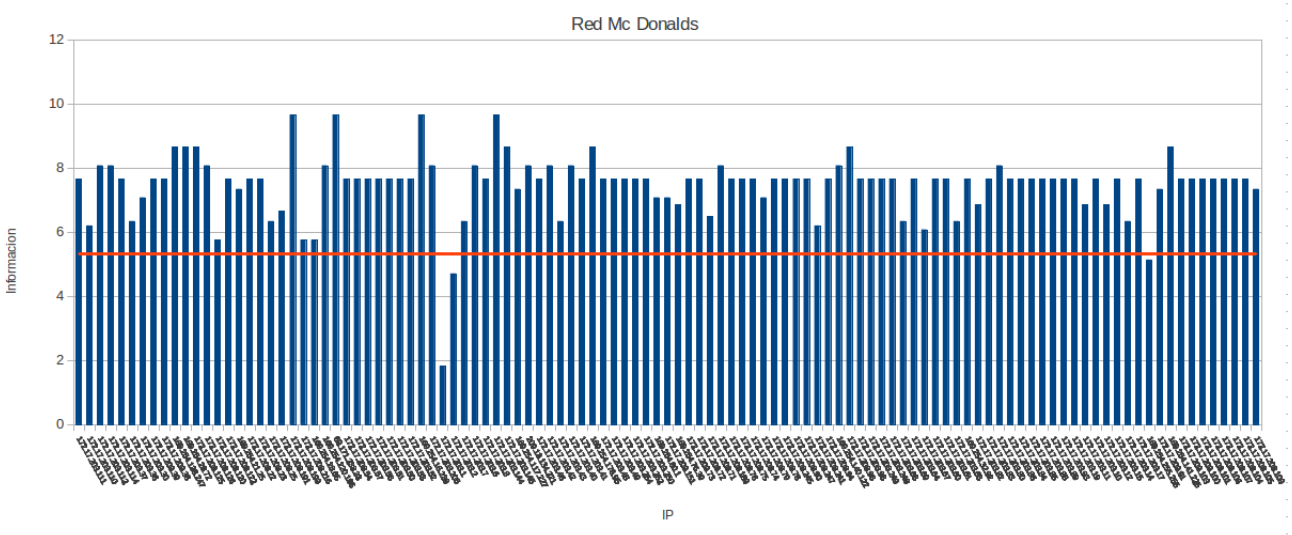


Figura 5: Cantidad de información de cada IP como destinatario, red McDonald's

Lamentablemente, por la gran cantidad de IPs que participaron en la comunicación, en el gráfico no se pueden ver con claridad las IPs pero en la sección de discusión se hablará de las IPs que tienen relevancia para nosotros.

Además se hizo un grafo reducido de la comunicación entre las IPs de la red, para ver algunos aspectos interesantes de la red. El grafo tiene el mismo estilo que el confeccionado para la red del Alto Palermo: cada nodo representa una IP y existe un eje dirigido de una IP hacia otra si hubo un envío.

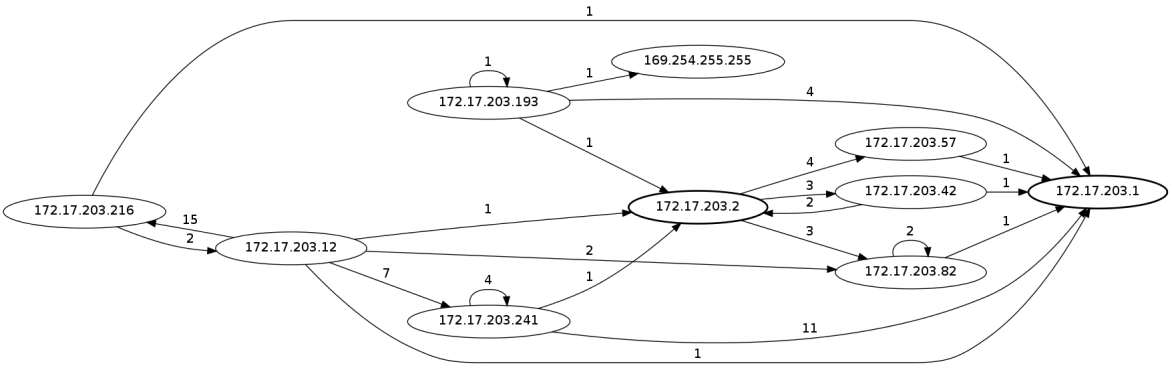


Figura 6: Grafo reducido de paquetes para la red del McDonald's

Un aspecto interesante que se ve rápido en este grafo es que existen pedidos ARP donde el emisor y el destinatario son la misma IP (por ejemplo, `172.17.203.241` o `172.17.203.82`). Además se observa que el grueso de pedidos va a parar al dispositivo que tiene asignada la IP `172.17.203.1`, es decir, el router que oficia de *gateway*.

5. Discusión

En esta sección se presenta un análisis un poco más profundo y las deducciones que se pueden extraer de los experimentos realizados. Se separa el análisis por cada red escuchada.

5.1. Red del Alto Palermo

En esta red se observa una gran interacción con el router, cuya IP es `172.17.0.1`. Esto sucede porque el router es el dispositivo por el que pasan en general la mayoría de los pedidos. Un aspecto interesante que se ve en esta comunicación es que no parece haber un pedido fuera de la red, es decir, a otro dominio. Lo que se puede deducir es que el dominio de esta red parece ser `172.17.0.0 /16` donde la IP del router es `172.17.0.1`, el cuál se comunica con el resto de los dispositivos bajo este dominio.

5.2. Red de Subway

Lo primero que se ve es la baja cantidad de IPs que interactuan en la comunicación. Además se observa que la interacción se realiza en una red privada del dominio `192.168.0.0 /16` que se suele usar para redes domésticas. Naturalmente la IP del router en este tipo de red es `192.168.0.1`.

De los datos obtenidos, se observó que casi todos los pedidos tenían como destinatario la IP del router. Algo que se puede ver en el histograma realizado para esta red es que la IP que menos información tiene es `192.168.1.20`, lo que significa que es la que más pedidos ARP efectuó.

5.3. Red de Casa

La primera observación que se puede realizar es que a pesar de contar unicamente con 5 diferentes dispositivos con acceso a la red (incluyendo el router) se pueden observar un total de 13 diferentes IPs. En segunda instancia se puede detectar fácilmente la linea de la entropía en un valor aproximado de 2,1 y así también la única IP que aporta menor información que la entropía a la IP del router en valor de `192.168.0.1` el cual da salida hacia internet para los dispositivos de la red. Notar que el dominio de la IP del router vuelve hacer la red doméstica `192.168.0.0 /16`, lo que es natural porque el experimento fue realizado en un departamento.

5.4. Red de McDonald's

Esta red es la que más datos aportó al análisis, fue claramente en la que más pedidos ARP se hicieron entre distintos nodos. Esto hace que tenga algunos aspectos interesantes para observar, en comparación con las otras redes analizadas. A diferencia de las anteriores redes se puede destacar que al ser una red de mayor incidencia (contar con una cantidad alta de dispositivos en la comunicación), la diferencia de información entre la IP de menor información (el router) y la entropía de la fuente es mayor.

Lo que no se observa bien en el histograma es que la mayoría de las IPs son del estilo `172.17.203.X`, lo que nos induce a pensar que el dominio de la red es `172.17.203.0 \24` y que la mayor interacción se produce en ese dominio, en donde la mayoría de los pedidos ARP tienen como destinatario la IP `172.17.203.1` que se deduce que la IP correspondiente al router de la LAN (es la que menos información aporta en la fuente graficada en el histograma, es decir, es la barra más pequeña).

Otro aspecto interesante que se observó es la gran cantidad de veces que figuraba `0.0.0.0` como IP emisora. Este es un pedido ARP gratuito que puede ser utilizado por cualquier dispositivo para verificar que determinada IP no esté siendo usada (la IP destino del pedido). La aparición de esta IP tiene relación con el siguiente hecho: la IP destino de estos pedidos siempre pertenecían al dominio `169.254.0.0 /16`, usada como *broadcast* de *link local*. Se usa para comunicación entre hosts en un sólo *link*. Los hosts obtienen direcciones de este dominio mediante una autoconfiguración, por ejemplo, en situaciones donde el servidor DHCP (*Dynamic Host Configuration Protocol*) no se encuentra. Es decir, cuando una interfaz pierde la conexión o es activada, si no puede recibir una IP con el protocolo DHCP se le asigna una IP del dominio `169.254.0.0 /16`. Por lo tanto, con el pedido de parte de `0.0.0.0` se quiere fijar si la IP que le fue asignada no está siendo usada en la red. El último comentario con este dominio, que también está relacionado con el pedido gratuito, es el siguiente: se observó un pedido ARP donde la IP emisora era la misma que la destinataria por cada IP del dominio `169.254.0.0 /16`. Esta particularidad la vimos en el taller; cuando se levanta una interfaz muchos sistemas envían automáticamente un pedido ARP gratuito con esta característica. De esta manera se pueden detectar IPs duplicadas (en caso de recibir respuesta) y actualiza la caché ARP de los demás hosts de la red.

6. Conclusiones

Se pudo observar que capturar los paquetes ARP en distintas redes locales da bastante información sobre la red. Es decir, con un análisis estádistico sobre los paquetes ARP *who-has* tomando como referencia las fuentes de información que se pueden definir a partir de las IPs que emiten mensajes y las IPs detinatarias, se puede por ejemplo deducir cuál es el router de la red. Otra manera de verlo es que si ya se sabe cuál es la IP del router, los experimentos sirven para ver que efectivamente el router es el dispositivos más solicitado, como se puede ver en la experimentación que se realizó en el departamento de un miembro del equipo.

Se ejecutó la herramienta implementada en distintas redes, lo que supuso distintos resultados. Lo interesante de esto es detectar nodos distinguidos o un comportamiento especial que hizo que tuviéramos que averiguar ciertas cosas, como sucedió en las capturas de la red del McDonald's en donde aparece una forma de asignar IPs cuando se levantan interfaces. Asimismo, con los histogramas hechos de las fuentes correspondientes se puede ver el grado de actividad de ciertos dispositivos en la red (por lo menos en el momento en que fue realizado el experimento; para mayor precisión se debería realizar una captura más amplia y de más duración).

Como última observación cabe destacar que no se debió participar 'agresivamente' en la red. Es decir, la escucha de los paquetes ARP fue pasiva. Sólo con recopilar información durante un tiempo determinado se puede deducir con bastante confiabilidad qué dispositivo hace de servidor para el resto de los nodos de la red.

7. Referencias

- RFC 826 (An Ethernet Address Resolution Protocol) `http://tools.ietf.org/html/rfc826`, David C. Plummer, 1982
- RFC 5735 (Special-Use IPv4 Addresses) `http://tools.ietf.org/html/rfc3330`, IANA (*Internet Assigned Numbers Authority*), 2002
- `http://www.secdev.org/projects/scapy/`
- `http://www.wireshark.org`
- `http://www.tcpdump.org`