Departamento de Computación,
Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires

Trabajo Práctico 1

Teoría de las Comunicaciones

Primer Cuatrimestre de 2014

Apellido y Nombre	LU	E-mail
Delgado, Alejandro N.	601/11	nahueldelgado@gmail.com
Lovisolo, Leandro	645/11	leandro@leandro.me
Petaccio, Lautaro José	443/11	lausuper@gmail.com

Índice

1. Introducción téorica

En este trabajo realizamos un análisis de redes mediante la captura de paquetes ARP.

Address Resolution Protocol (ARP) es un protocolo usado frecuentemente por las redes locales para conectar las capas 3 (capa de red) y 2 (capa de enlace) mediante la conversión o identificación de IP v4 con direcciones físicas MAC.

Existen dos tipos de paquetes posibles en el protocolo: paquetes de petición y de respuesta.

- Los paquetes de petición (*who-has*) son enviados mayormente en forma de broadcast con el objetivo de poder localizar la dirección MAC a la cuál le pertenece una IP conocida.
- Los paquetes de respuesta (*is-at*) son enviados de manera uni-cast ya que se utilizan para responder a la máquina que realizó una petición con anterioridad.

La estructura de los paquetes ARP es simple, consiste principalmente de los siguientes campos:

- Operación - Especifica la operación que el emisor está realizando: 1 para petición, 2 para responder
- Dirección MAC del emisor
- Dirección IP del emisior
- Dirección MAC del destinatario - Este campo se ignora en las peticiones
- Dirección IP del destinatario

A continuación se describe un ejemplo de uso típico observado en la práctica.

> Una máquina en una red quiere mandarle un paquete de datos a otra máquina en la misma red. Para esto, la máquina emisora busca en su tabla local, la dirección MAC asociada a la dirección IP a la cuál quiere mandar el paquete. Si no la encuentra, realiza el broadcasts de la petición ARP, la cual llegará eventualmente, si se encuentra conectada, a la máquina destino. La máquina destino recibirá la petición y la responderá de manera uni-cast hacia la máquina que realizó la petición, poniendo en el paquete su dirección MAC para que la máquina destino de la respuesta pueda conocer la dirección MAC que necesitaba.

El análisis de la red consiste en reconocer su topología en base al nivel de información que proveen las diferentes IP, como fuente y como destino, tomando a las IP como símbolos y estimado su probabilidad de aparición con su frecuencia muestral.

2. Desarrollo

Implementamos, para nuestro análisis, un *sniffer*[1] o monitor de paquetes, con el objetivo de poder analizar los paquetes ARP siendo enviados vía broadcast por el medio utilizado. Esta implementación se realizó en Python y utiliza la biblioteca Scapy [2] para la captura de paquetes, provista por la cátedra.

Los medios utilizados para la captura de paquetes fueron tres redes Wi-Fi, de acceso público y de frecuente utilización.

Identificamos las redes según su SSID:

- Alto Palermo
- McDonald's
- Starbucks

[1] `arpsniffer.py`
[2] `http://www.secdev.org/projects/scapy`

La captura de paquetes ARP consistió en el monitoreo de los paquetes ARP durante media hora aproximadamente para cada red Wi-Fi. Se almacenaron los campos principales de cada paquete para luego realizar el análisis estadístico.

Habiendo obtenido las capturas de las redes, nos proponemos identificar el o los routers que actúan como *gateway* de cada red analizada. Para esto, calculamos dos tipos de frecuencia muestral para cada IP, la primera en relación a cuántos paquetes *who-has* la tienen como emisor y la segunda, cuántos paquetes *who-has* la tienen como destino. De este análisis, conseguimos la frecuencia muestral de cada IP como emisor y como destino.

Para cada red consideramos las fuentes de información S_{src} y S_{dst} tales que sus símbolos son las direcciones IP que aparecen como origen y destino en los paquetes ARP *who-has*, respectivamente. Para cada fuente, calculamos la información de cada símbolo[3].

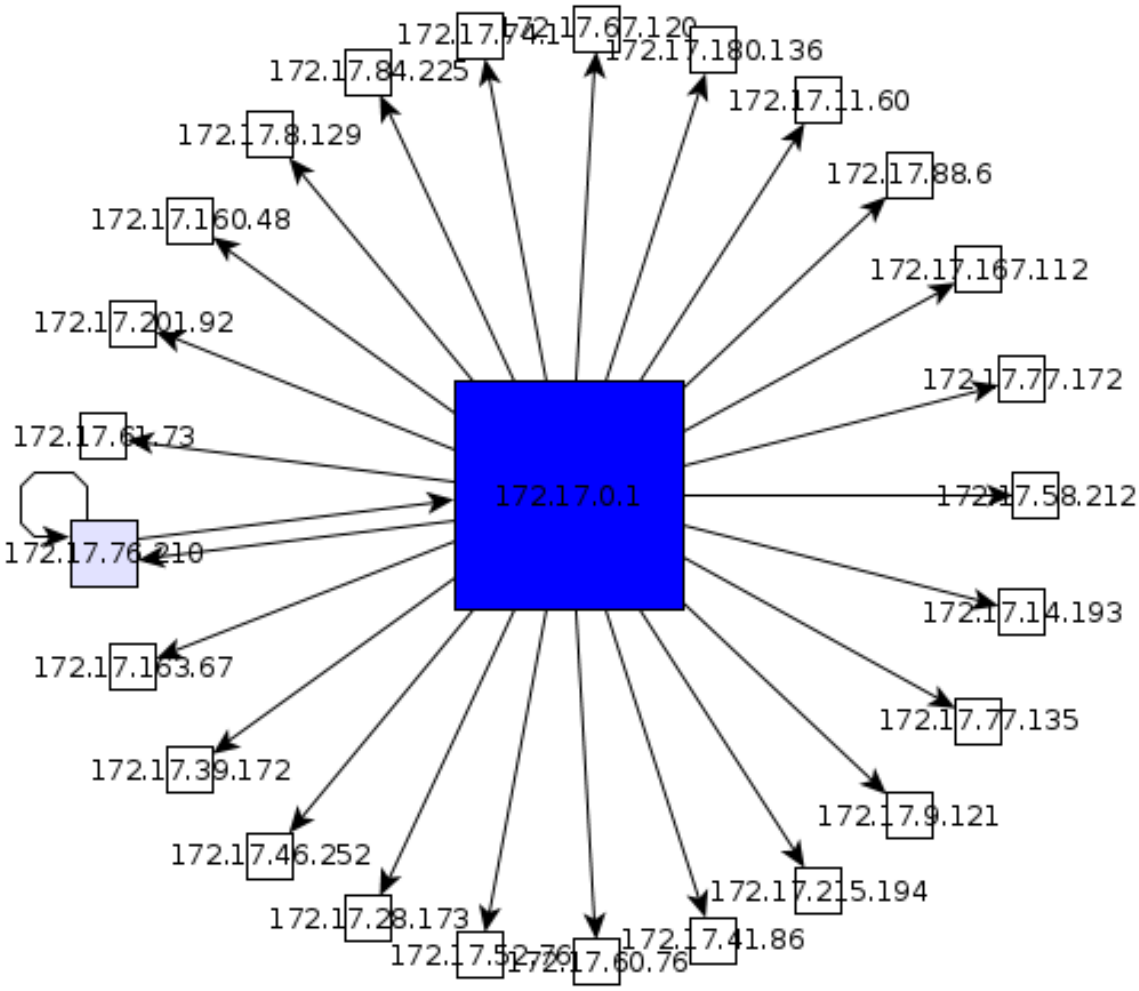
A continuación, realizamos la identificación de el o los routers en la red utilizando la información de cada IP para ambas fuentes S_{src} y S_{dst}. Según suponemos, las IP asociadas a los routers deberían poseer la menor cantidad de información debido a que su frecuencia de recepción e incluso emisión de paquetes ARP debería ser la más alta. Esta suposición se basa en que el router será el dispositivo a los que todos los equipos querrán comunicarse (para poder tener acceso a internet), y mantendrán su ubicación actualizada para poder realizar la comunicación.

Para obtener una visualización de la red con la que se está trabajando, graficamos las comunicaciones en la red en forma de grafos dirigidos, utilizando un script [4] en Python para representar los grafos en formato Trivial Graph Format[5] que luego graficamos con el software yEd[6]. Tomamos como nodos los diferentes IP de ésta y como aristas dirigidas la existencia de un paquete ARP con una fuente y un destino específico.

[3] Es decir, cada dirección IP asociada a esa fuente de información.
[4] `tgfizar.py`
[5] `http://en.wikipedia.org/wiki/Trivial_Graph_Format`
[6] `http://www.yworks.com/en/products_yed_about.html`

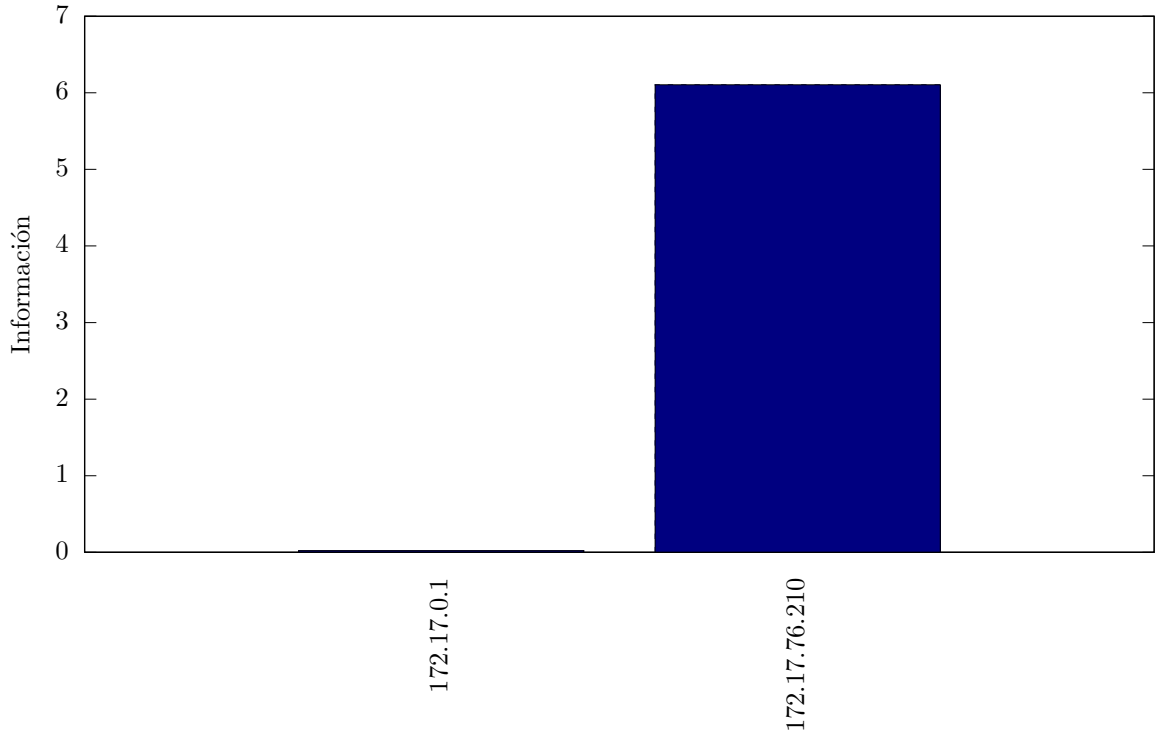
3. Resultados

3.1. Red *Alto Palermo*



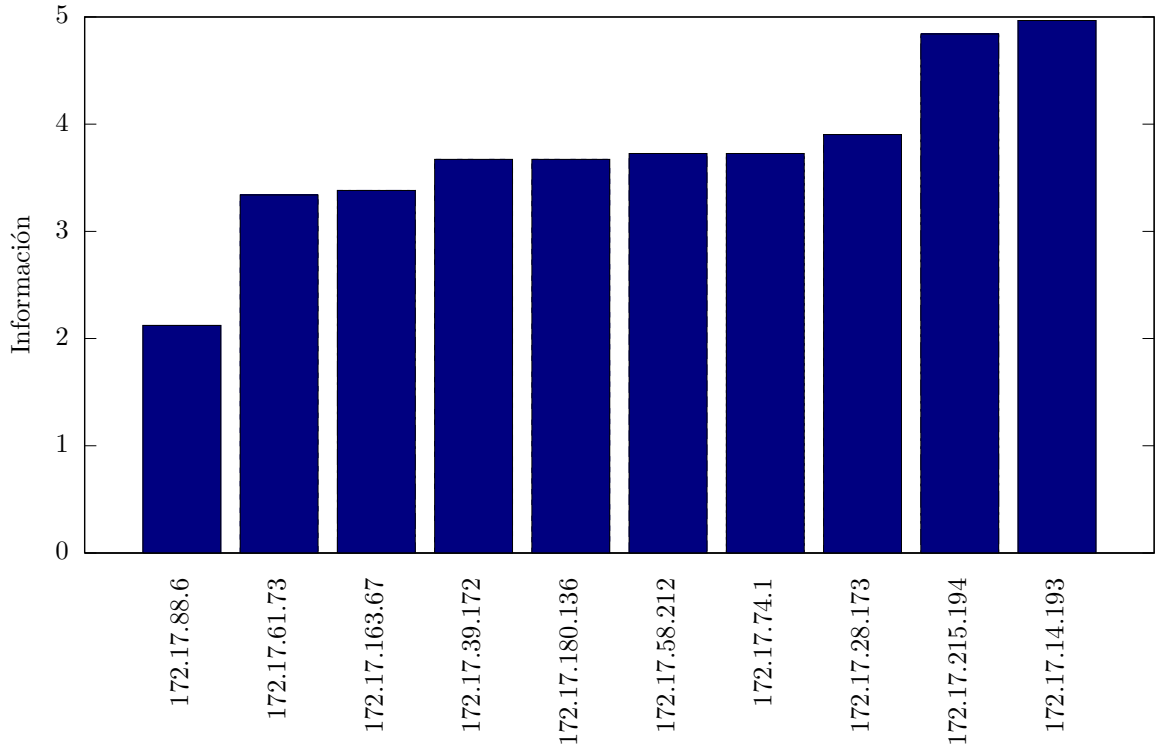
3.1.1. Información de los símbolos de la fuente de información S_{src}

Donde $S_{src} = \{$direcciones IP origen en paquetes ARP *who-has*$\}$.

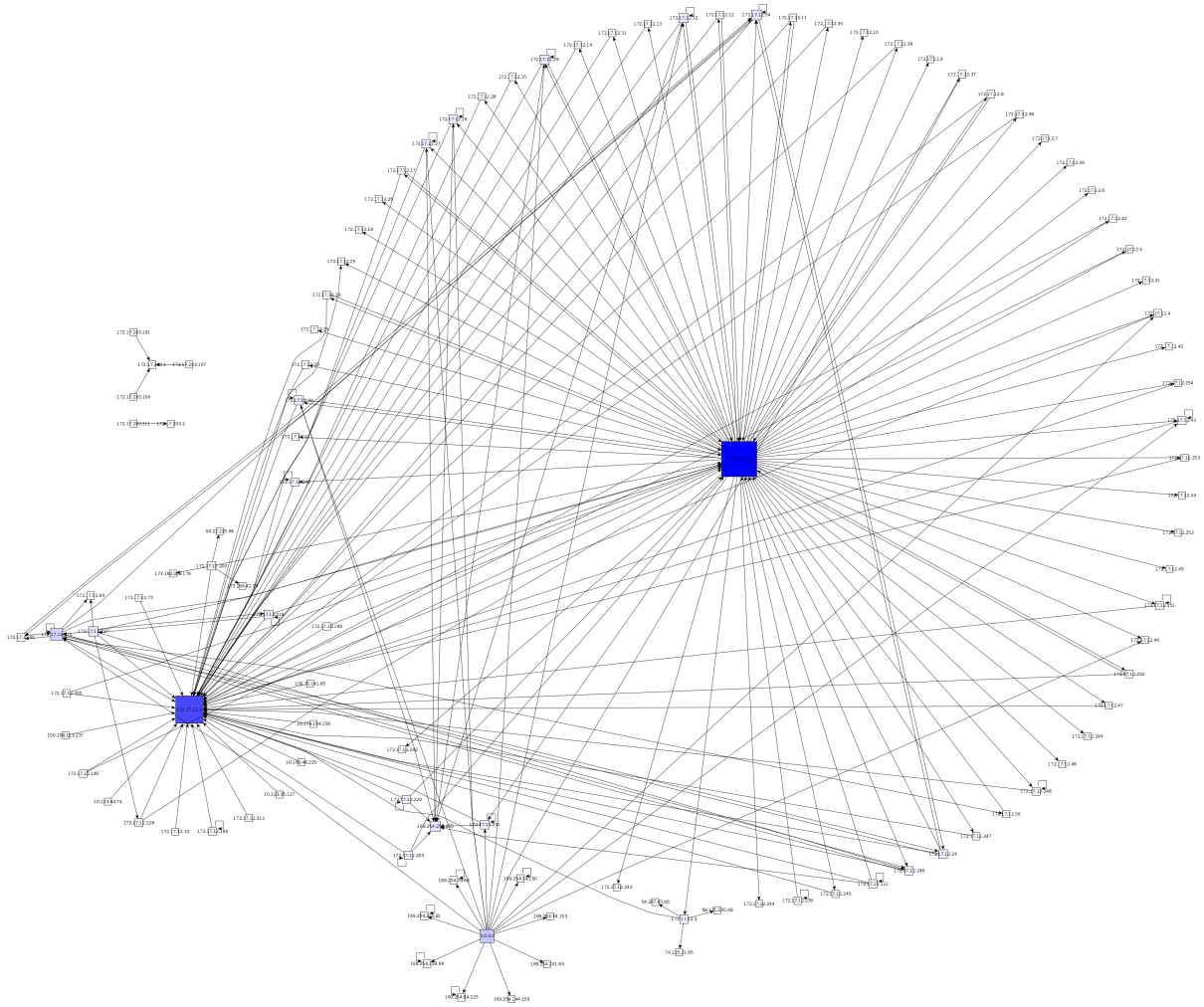


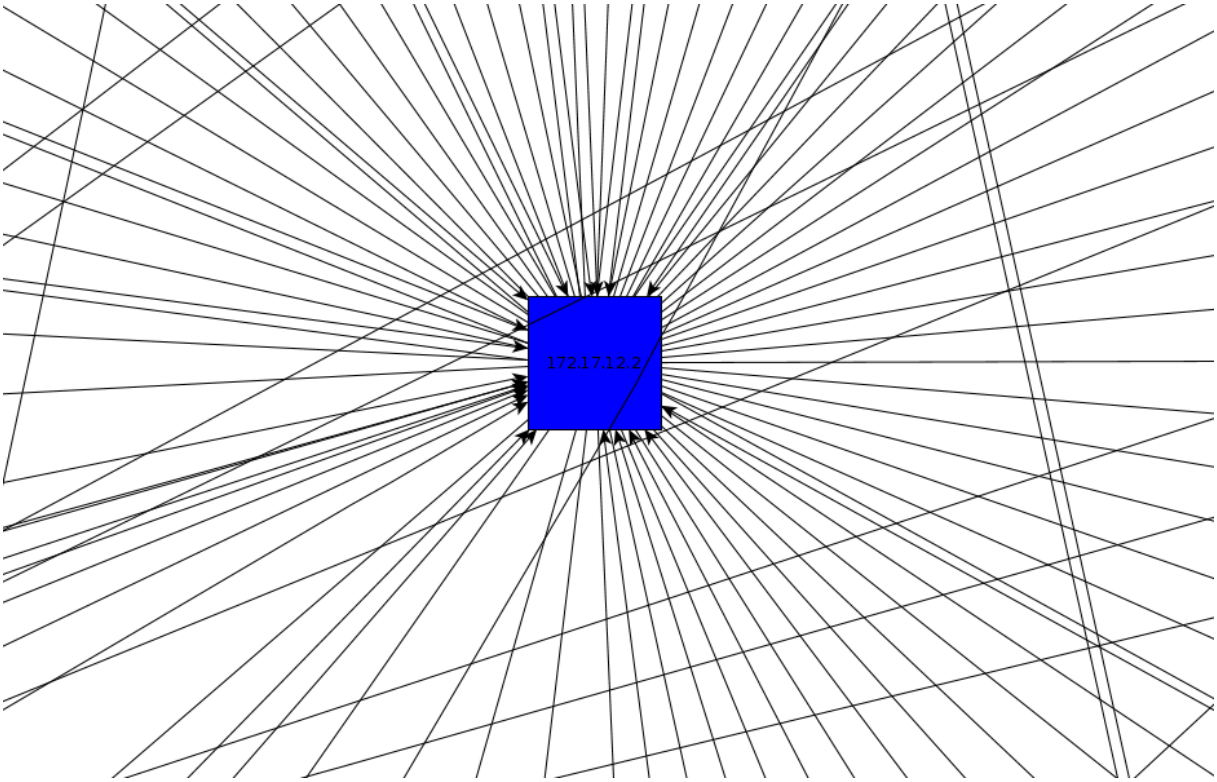
3.1.2. Información de los símbolos de la fuente de información S_{dst}

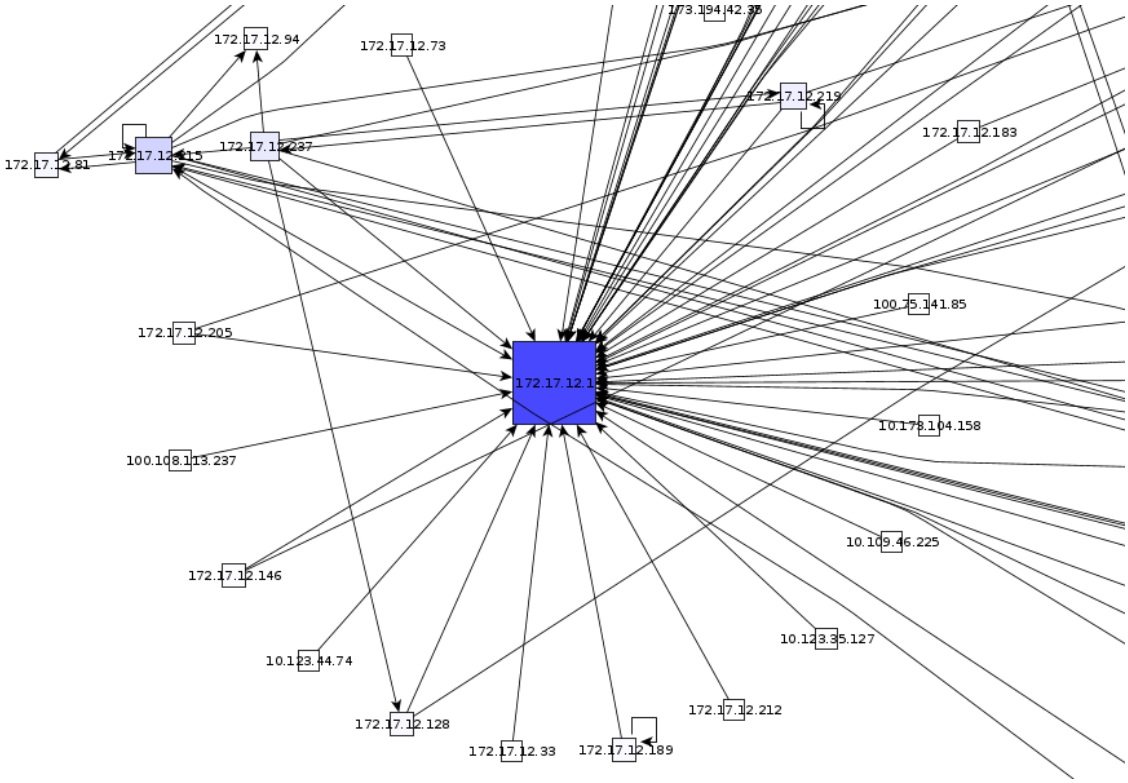
Donde $S_{dst} = \{$direcciones IP destino en paquetes ARP *who-has*$\}$.

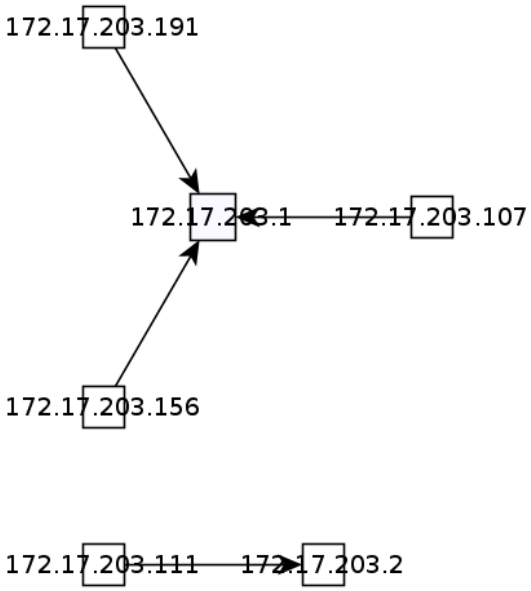


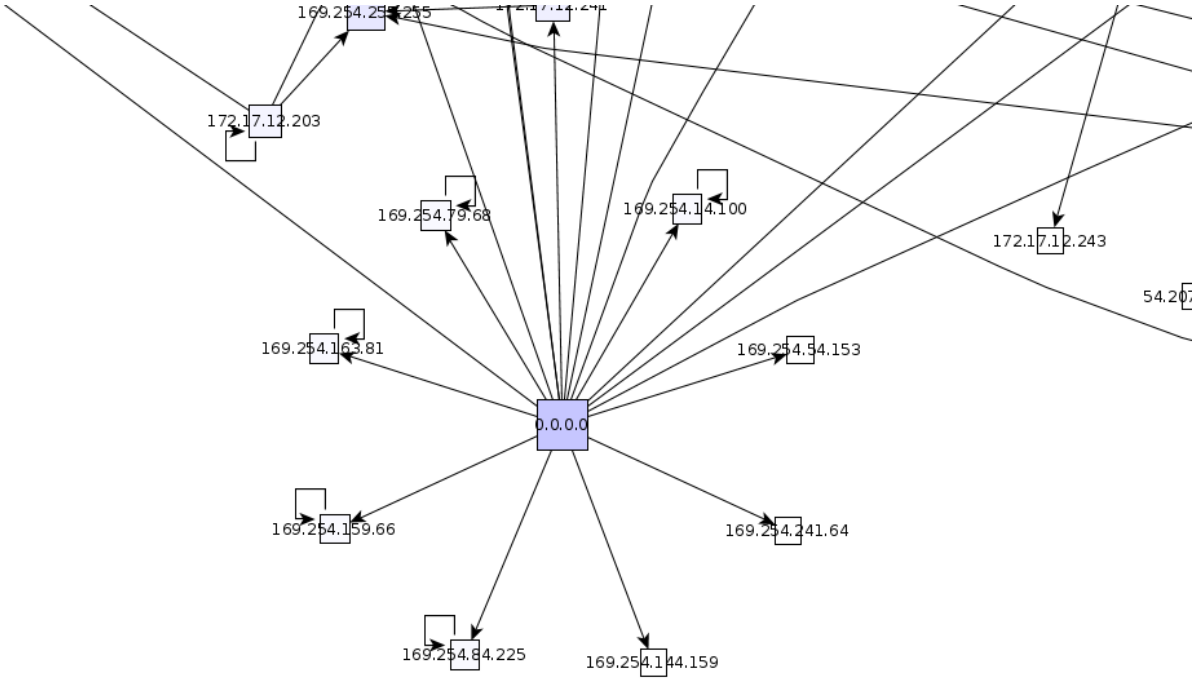
3.2. Red *McDonald's*




172.17.12.2

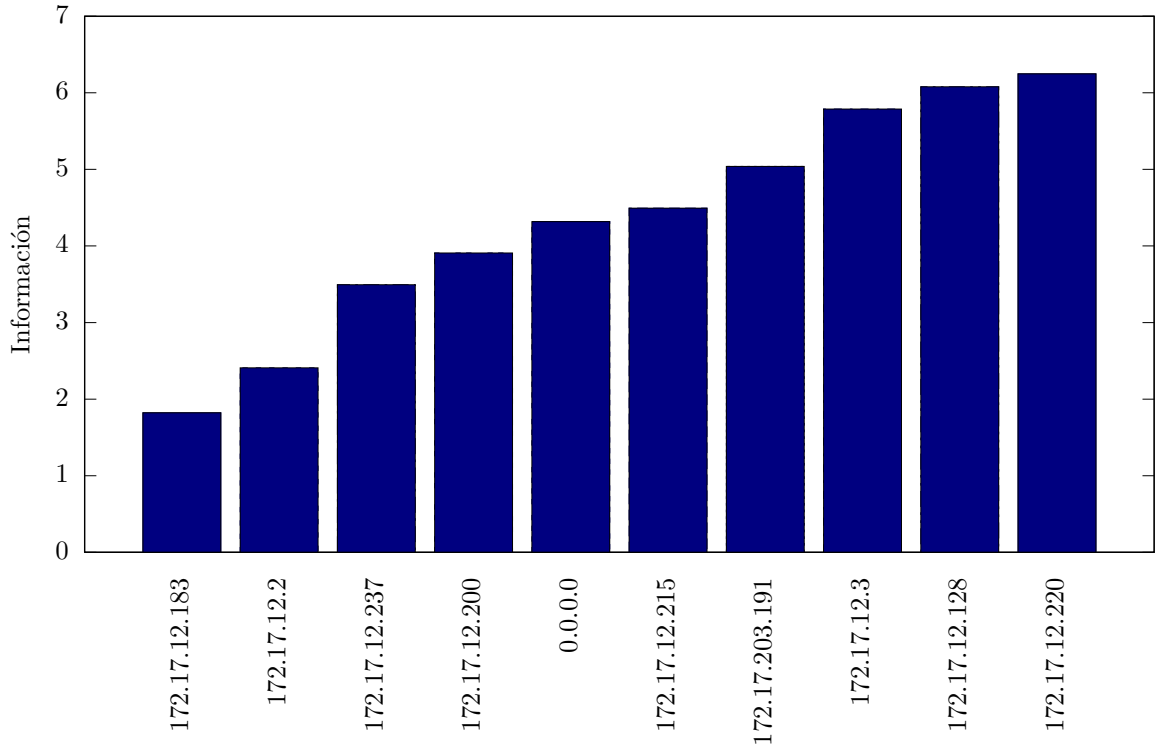

172.17.12.94
172.17.12.73
172.17.12.219
172.17.12.183
172.17.12.81
172.17.12.115
172.17.12.237
100.25.141.85
172.17.12.205
172.17.12.1
10.178.104.158
100.108.113.237
10.109.46.225
172.17.12.146
10.123.35.127
10.123.44.74
172.17.12.212
172.17.12.128
172.17.12.33
172.17.12.189





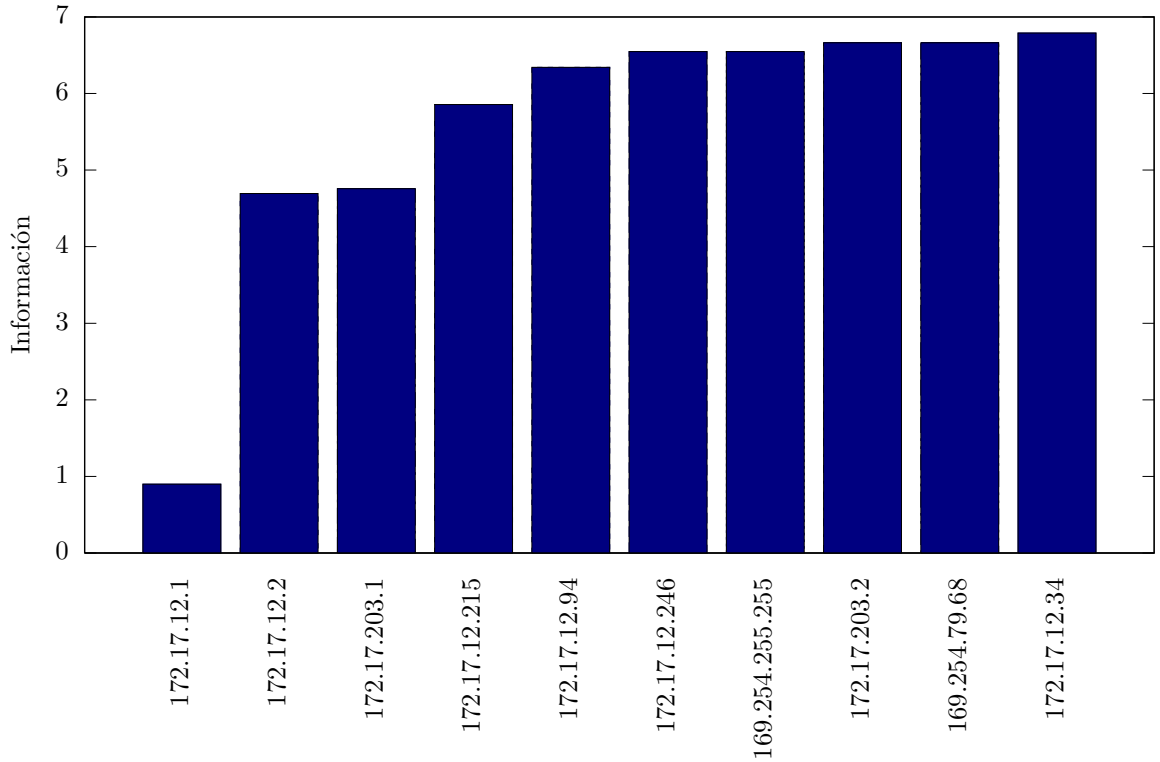
3.2.1. Información de los símbolos de la fuente de información S_{src}

Donde $S_{src} = \{$direcciones IP origen en paquetes ARP *who-has*$\}$.

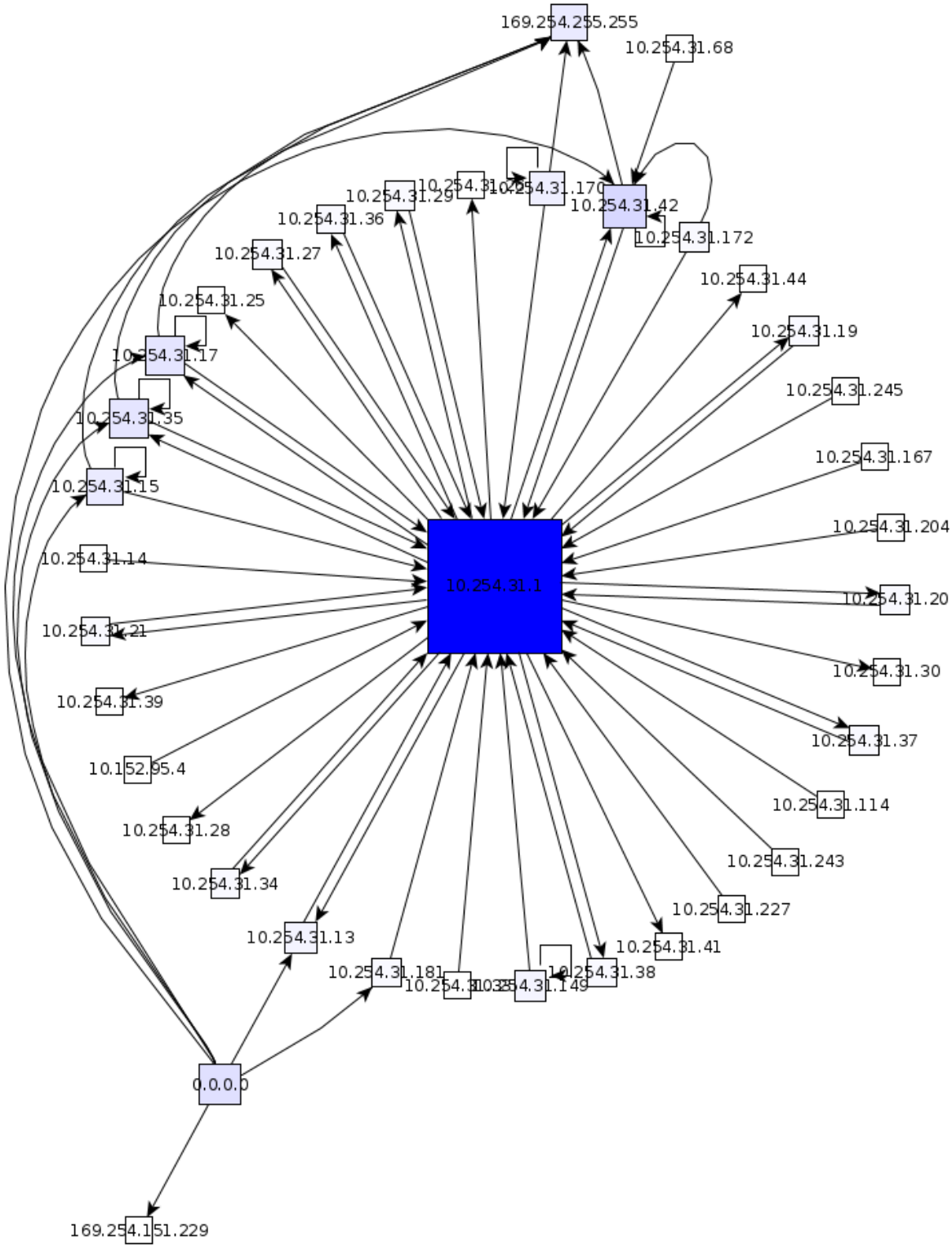


3.2.2. Información de los símbolos de la fuente de información S_{dst}

Donde $S_{dst} = \{$direcciones IP destino en paquetes ARP *who-has*$\}$.

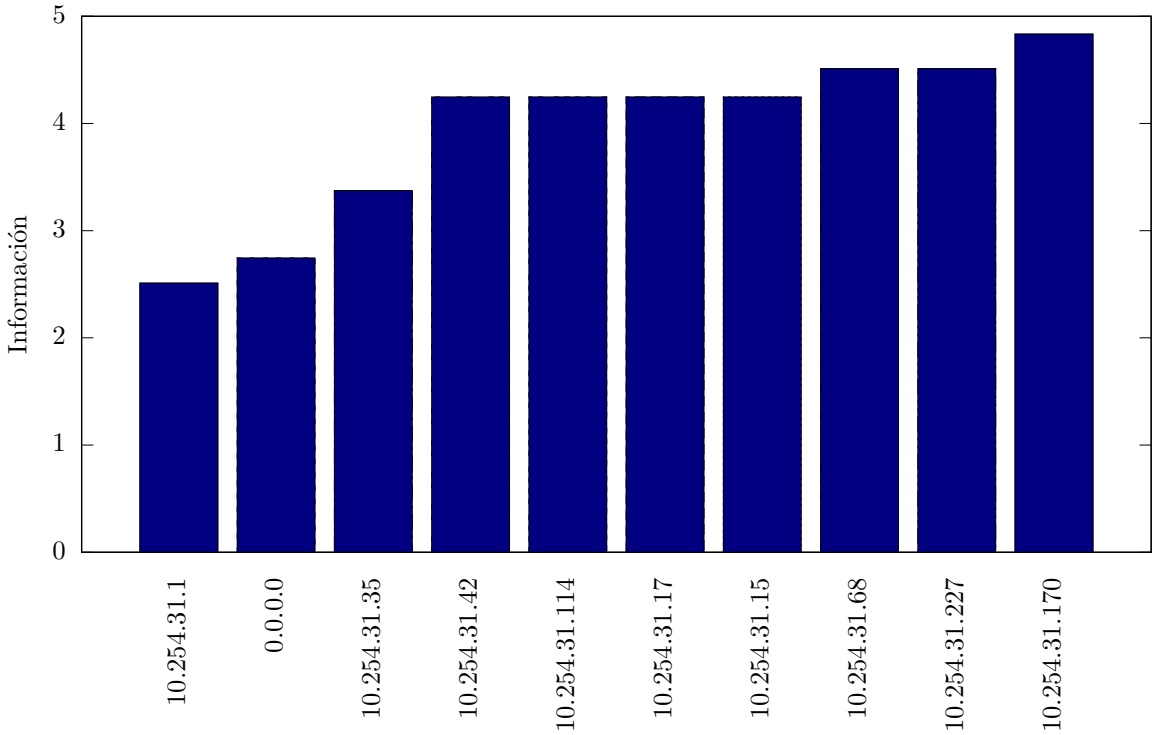


3.3. Red *Starbucks*



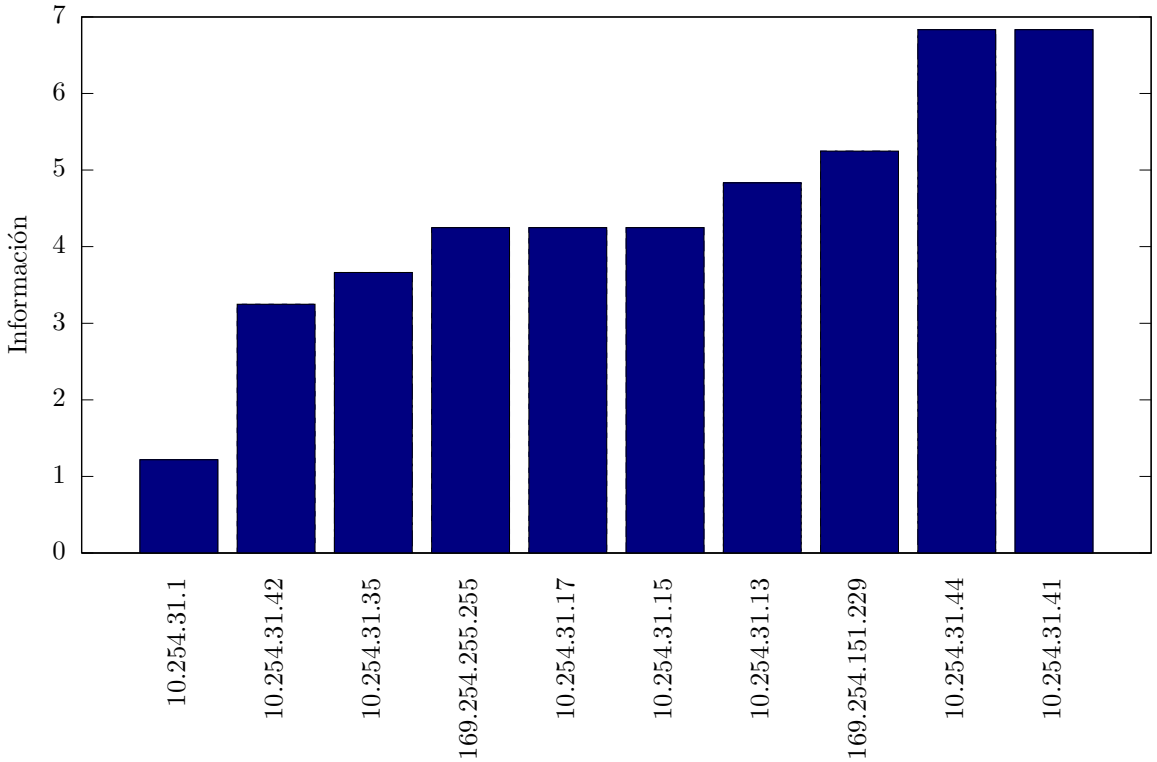
3.3.1. Información de los símbolos de la fuente de información S_{src}

Donde $S_{src} = \{$direcciones IP origen en paquetes ARP *who-has*$\}$.



3.3.2. Información de los símbolos de la fuente de información S_{dst}

Donde $S_{dst} = \{$direcciones IP destino en paquetes ARP *who-has*$\}$.



3.4. Estadísticas

3.4.1. Tamaño de las muestras y tiempos de captura

Red	Tamaño de la muestra	Tiempo de captura (minutos)
Alto Palermo	344	17
McDonald's	1216	28
Starbucks	114	45

3.4.2. Entropía

Red	Fuente de información modelada S	Entropía $H(S)$
Alto Palermo	S_{src}	0.109542
Alto Palermo	S_{dst}	3.738743
McDonald's	S_{src}	3.976417
McDonald's	S_{dst}	3.769457
Starbucks	S_{src}	4.142124
Starbucks	S_{dst}	3.271891

4. Discusión

Describimos a continuación nuestra interpretación de algunos fenómenos observados, de manera de facilitar la discusión de los resultados obtenidos.

4.1. Nodos que emiten paquetes ARP *who-has* hacia un único nodo destino

Dado que la captura de paquetes fue realizada sobre redes wireless públicas, asumimos que la mayoría de las personas conectadas a la red sólo desean obtener acceso a internet, con lo que sus dispositivos únicamente solicitan la dirección física del gateway que les asigna al conectarse a la red.

4.2. Nodos que emiten paquetes ARP *who-has* hacia muchos nodos destino

Teniendo en cuenta lo anterior, interpretamos que este escenario es generalmente producido por un router que intenta mantener actualizada su tabla ARP emitiendo paquetes ARP *who-has* periódicamente para cada dirección IP en dicha tabla, bajo la suposición que cada dirección IP es asignada a distintos hosts que se conectan y desconectan de la red a medida que transcurre el tiempo.

4.3. Nodos que no emiten ningún paquete ARP

Suponemos que estos nodos se conectaron a la red y resolvieron la dirección física del gateway antes de iniciar la captura. De acuerdo con lo dicho anteriormente, cada nodo recibe periódicamente un paquete ARP *who-has* emitido por el router al que está conecatdo. Este paquete incluye sus direcciones IP y física. Suponemos que cada nodo refresca su tabla ARP con esta información, lo que evita la necesidad de emitir paquetes ARP *who-has* para solicitar la dirección física del router ya que esa entrada en su tabla ARP nunca llega a caducar.

4.4. Fenómenos Destacables

4.4.1. Paquetes ARP con IP origen 0.0.0.0

Pudimos detectar en las redes algunos paquetes ARP *who-has* con IP origen 0.0.0.0. El motivo de uso de esta IP es el siguiente: Cuando un cliente se conecta a una red que posee un servidor DHCP y quiere recibir una IP de éste, manda una petición con su ID en forma de broadcast para que lo detecte el servidor. Una vez detectado por el servidor, éste manda una o varias ofertas de IP a ese ID. El cliente

eventualmente podría recibir la oferta, tomar uno de esos IP y extraer la dirección del router. Como el servidor realiza la misma operación con todos los demás clientes que pidan una IP, el cliente debe comprobar que la IP que eligió no la tiene otro cliente. Para esto, envía un paquete *who-has* con su MAC adress y la IP 0.0.0.0 como fuente para evitar confundir las ARP caches en otros hosts. Si el *who-has* es respondido, el cliente rechaza el IP elegido.

4.4.2. Direcciones en el rango 169.254.0.0/16

Entre los paquetes capturados detectamos varios que usaban el rango 169.254.0.0/16 como dirección IP origen o destino, el cual difiere con el rango utilizado para las IP asignadas por el servidor DHCP a los dispositivos de la red. Las direcciones en este rango se denominan *direcciones de enlace local*, y son direcciones reservadas. Un host puede eventualmente asignarse una IP libre (lo corrobora con ARP) de enlace local para poder acceder de forma básica a la red cuando todavía no se le ha asignado una IP válida, ya sea de forma manual o automática (DHCP).[7]

Esto le permite al host comunicarse con los otros dispositivos de la red, pero no con dispositivos externos a la misma.

4.4.3. Misma IP como origen y destino

Este escenario se presentó en muchas ocasiones, en todas las redes analizadas. Se trata de una forma de uso del protocolo llamada *Gratuitous ARP*[8], que existe mayormente para detectar conflictos de IP en la red, o para actualizar la información en las tablas de los vecinos.

Por ejemplo, si un host envía un Gratuitous ARP y recibe una respuesta, ya detectó un conflicto de IP, pues nadie debería responder a un request que tiene como destino el propio host.

5. Conclusión

[7] `http://tools.ietf.org/html/rfc3927`
[8] `http://wiki.wireshark.org/Gratuitous_ARP`