DEPARTAMENTO
DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

# TP1: Wiretapping

Teoría de las Comunicaciones
Segundo Cuatrimestre de 2021


| Integrante | LU | Correo electrónico |
|---|---|---|
| Christian Nahuel Rivera | 184/15 | christiannahuelrivera@gmail.com |
| Guido Rodriguez Celma | 374/19 | guido.rc98@gmail.com |
| Miguel Rodriguez | 57/19 | mmiguerodriguez@gmail.com |





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

# Índice

**1. Introducción** **2**

**2. Modelando tráfico** **2**

**3. Análisis de redes** **4**

3.1. Redes heterogéneas . . . . . 4

3.2. Servicios de streaming . . . . . 7

**4. Nodos distinguidos** **9**

**5. Conclusiones** **12**

# 1. Introducción

Los medios de acceso compartido (i.e.: Ethernet, WiFi) son un recurso ampliamente difundido para establecer enlaces de comunicación entre 2 o más computadoras. Los paquetes de datos que se transmiten por estos medios llegan a todos los dispositivos conectados al mismo y se usan las direcciones en los encabezados de los paquetes para poder enviar información de un dispositivo a otro específicamente, comúnmente llamado de tipo Unicast. Además, existen varios protocolos que hacen uso de comunicaciones de tipo Broadcast, por ejemplo ARP y DHCP. Estos paquetes de datos pueden ser capturados por cualquier dispositivos que se encuentre conectado al medio y pueden ser analizados, por lo menos hasta la última capa de comunicación que no se encuentre encriptada.

En el presente trabajo nuestro objetivo será modelar los mensajes intercambiados en distintas redes y crear herramientas para poder analizar las métricas vistas en la materia basándonos en la teoría de la información de Shannon.

Capturaremos tramas de la capa de enlace, que es la segunda capa en el modelo OSI. Esta es la encargada de transmitir datos entre nodos dentro de un segmento de red a través de la capa física. La capa de enlace provee la parte funcional para transmitir datos entre redes y además detectar y potencialmente corregir errores que pueden ocurrir en la capa física.

# 2. Modelando tráfico

Una fuente de información de memoria nula $S$ es una fuente que emite distintos símbolos $s$, en donde cada uno tiene una probabilidad asociada y además son estadísticamente independientes. Cada símbolo $s$ aporta una cantidad de información $I(s) = -\log(P(s))$ donde $P(s)$ es la probabilidad de ocurrencia del símbolo $s$ dentro de la fuente $S$.

A partir de la información que nos brindan los símbolos de una fuente podemos calcular la entropía de la misma, la cual nos da una medida de la incertidumbre de una fuente de información. La entropía también puede verse como la cantidad promedio de información que contienen los símbolos usados. El valor de la entropía de una fuente de información esta dada por la formula $\sum_{s \in S} P(s)I(s)$ y lo usaremos para ver cuales son los símbolos distinguidos que aportan mayor información para las comunicaciones en el marco de este protocolo.

Utilizando la herramienta provista por la cátedra realizamos capturas de los paquetes que circulan en las redes privadas de los integrantes del trabajo. Luego, a partir de la información obtenida por cada captura, calculamos las métricas detalladas anteriormente. El modelo usa una fuente S1 cuyos símbolos siguen el formato: `| Protocolo | Tipo |` donde el tipo se refiere al método de envío (Broadcast o Unicast).

**Red 1**
Para la captura en un red a la cual estábamos conectados por Ethernet, se utilizaron los datos de 50.000 tramas. Podemos ver en la tabla 1 la cantidad de información obtenida por cada paquete según su símbolo. Observamos que los paquetes de tipo `IPv4/UNICAST` suelen proporcionar menos información que los demás ya que su ocurrencia es mucho mayor por el hecho de ser utilizados para comunicar datos de usuario entre distintos dispositivos, a diferencia de paquetes de tipo `ARP/BROADCAST` o `ARP/UNICAST` que son protocolos de control.

| **Tipo de paquete** | **Probabilidad de ocurrencia** | **Información de cada simbolo** |
|---|---|---|
| IPv4/UNICAST | 0.99720 | 1 |
| IPv6/UNICAST | 0.00106 | 10 |
| ARP/BROADCAST | 0.00066 | 11 |
| IPv4/BROADCAST | 0.00058 | 11 |
| ARP/UNICAST | 0.00048 | 12 |
| LLDP/UNICAST | 0.00002 | 16 |

Cuadro 1: Captura de red Ethernet

La entropía obtenida para esta fuente de información fue de 1.02752.

**Red 2**
En esta red nuevamente capturamos los datos de 50.000 tramas, a diferencia de la anterior, esta era un red de tipo Wi-Fi. Los resultados obtenidos fueron los siguientes:

| **Tipo de paquete** | **Probabilidad de ocurrencia** | **Información de cada simbolo** |
|---|---|---|
| IPv4/UNICAST | 0.85408 | 1 |
| IPv6/UNICAST | 0.14572 | 3 |
| ARP/UNICAST | 0.00012 | 14 |
| 1905.1IEEE/UNICAST | 0.00004 | 15 |
| LLDP/UNICAST | 0.00004 | 15 |

Cuadro 2: Captura de red Wi-Fi

La entropía obtenida para esta fuente de información fue de 1.29412.

Al ser esta una red Wi-Fi observamos la aparición de paquetes de tipo `IEEE 1905.1`, que no circulan en la red Ethernet anterior, además esta red tiene un porcentaje mucho mayor de paquetes que utilizan el nuevo protocolo `IPv6`. Otra observación interesante es que todos los mensajes capturados fueron de tipo `UNICAST`, es decir que fueron enviados específicamente a un único dispositivo.

**Red 3**
Por tercera vez capturamos los datos de 50.000 tramas, desde otra red Wi-Fi. Es una red privada que tiene varias computadoras conectadas. Los resultados obtenidos fueron los siguientes:

| Tipo de paquete | Probabilidad de ocurrencia | Información de cada símbolo |
|---|---|---|
| IPv6/UNICAST | 0.89272 | 1 |
| IPv4/UNICAST | 0.10566 | 4 |
| ARP/BROADCAST | 0.00096 | 11 |
| ARP/UNICAST | 0.00054 | 11 |
| IPv4/BROADCAST | 0.00012 | 14 |

Cuadro 3: Captura de otra red Wi-Fi

La entropía obtenida para esta fuente de información fue de 1.33354.

La red estaba siendo usada principalmente por uno de los dispositivos conectados que estaba consumiendo datos de un servicio de `streaming`. El resto de los dispositivos solo estaba conectado a la red esperando actualizaciones. Podemos ver claramente como el consumo de video por uno de los dispositivos acapara el tráfico de la red. A diferencia de la otra red Wi-Fi, en esta obtuvimos algunos frames de `BROADCAST` pero son menores al 0.1 % del tráfico. Con los datos obtenidos en las tres redes pudimos observar que los frames de `BROADCAST` tienen un uso muy bajo lo cual es de esperarse ya que si ocurrieran seguido ralentizaría mucho al consumo de datos por el hecho de que se transmiten a todos los dispositivos de la red. Su uso debe ser mínimo para mantener actualizadas las tablas de direcciones y los dispositivos conectados.

# 3. Análisis de redes

En la siguiente sección se detallaran experimentos realizados con la herramienta implementada anteriormente aplicado sobre distintas redes.

## 3.1. Redes heterogéneas

Buscamos determinar qué diferencias o similitudes tienen distintas redes LAN con respecto al contenido de su tráfico, para este fin intentamos analizar redes lo más diversas posibles. Aplicamos el modelo de fuente $S1$ detallado anteriormente, donde los símbolos siguen el formato `Protocolo/Tipo` y los paquetes capturados son aquellos de tipo Ethernet.

**Red pública - cafetería:** En este caso decidimos analizar una red en la que constantemente se conectan y desconectan los hosts, para esto capturamos paquetes de la red Wi-Fi de un local de café en un horario concurrido. Usando el modelo ya presentado tomamos una muestra de 30.000 tramas.

```
Tipo de paquete    : Probabilidad de ocurrencia
 'IPv4/UNICAST'    : 0.91707
 'ARP/BROADCAST'   : 0.07200
 'IPv4/BROADCAST'  : 0.00583
 'ARP/UNICAST'     : 0.00293
 'IPv6/UNICAST'    : 0.00218

Entropía de la fuente: 1.29766
```

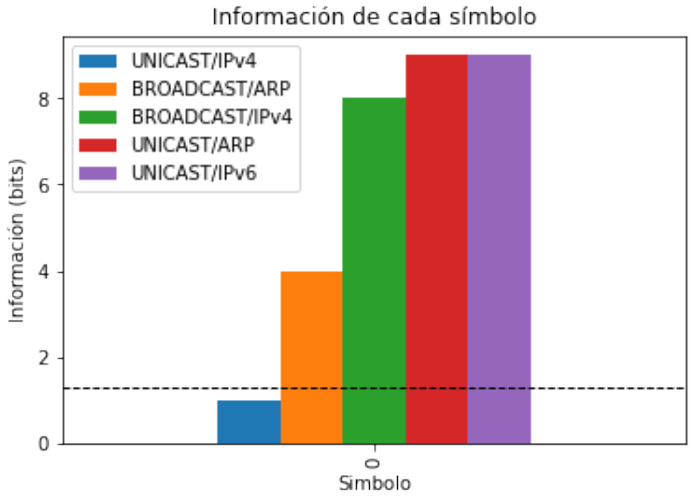


Figura 1: Red WLAN cafetería

**Red privada - doméstica:** Esta es la red de una casa de 4 inquilinos, la captura fue tomada en un horario donde todos estuvieran presentes usando la red para distintos propósitos. Nuevamente el tamaño elegido fue de 30.000 tramas.

```
Tipo de paquete          : Probabilidad de ocurrencia
 'IPv6/UNICAST'          : 0.71015
 'IPv4/UNICAST'          : 0.28229
 'ARP/UNICAST'           : 0.00375
 '1905.1IEEE/UNICAST'    : 0.00218
 'LLDP/UNICAST'          : 0.00171
 'IPv4/BROADCAST'        : 0.00032

Entropía de la fuente: 1.34724
```

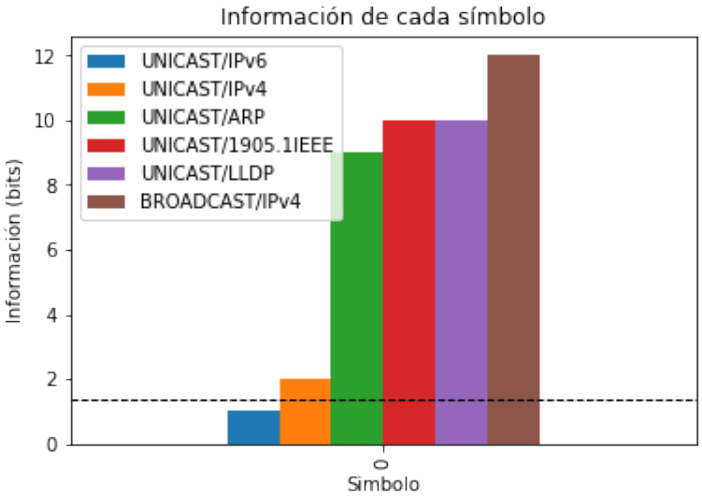


Figura 2: Red WiFi doméstica

**Red privada - PyME:** Esta es la red de una pequeña empresa que dispone de varias computadoras. La captura fue tomada en una de estas, conectada a la red mediante una interfaz Ethernet, en el horario de trabajo. Nuevamente el tamaño elegido fue de 30.000 tramas.

```
Tipo de paquete          : Probabilidad de ocurrencia
 'IPv4/UNICAST'          : 0.62302
 'IPv6/UNICAST'          : 0.36669
 'ARP/BROADCAST'         : 0.00368
 'LLDP/UNICAST'          : 0.00319
 'ARP/UNICAST'           : 0.00314

Entropía de la fuente: 1.44976
```

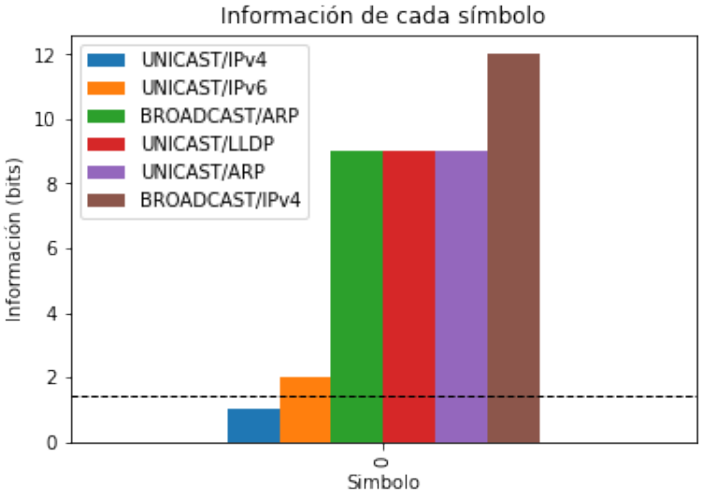


Figura 3: Red LAN PYME

A partir de los datos obtenidos observamos que en la red de la cafetería hay una presencia mucho mayor de paquetes de tipo `BROADCAST`, en particular del protocolo `ARP`, respecto de las otras dos redes analizadas. Creemos que esto se debe a que, al haber dispositivos nuevos constantemente incorporándose a la red, el o los routers de la misma deban en consecuencia reaprender la topología de la red cada vez, y para esto se valen del protocolo de Address Resolution. En las demás redes, el tráfico de tipo `BROADCAST` es despreciable respecto al `UNICAST`.

En ningún caso la entropía de la red alcanza su máximo, ya que las probabilidades de cada símbolo observado son en general dispares y no guardan relación entre sí (la entropía máxima se alcanza cuando todos los símbolos son equiprobables). No observamos una relación clara entre la entropía de la red y alguna característica de la misma, aunque la entropía de la red de la cafetería es algo menor que las demás. Esto podría deberse a la diferencia muy marcada entre cantidad de paquetes de tipo `IPv4/UNICAST` y todos los demás en la misma.

En todas las redes observamos en mayor o menor medida paquetes de datos y de control. Aquellos de transporte de datos de usuario son:

- `IPv4/IPv6`: Funcionan en la capa de red y su tarea principal es transferir bloques de datos desde un host fuente hasta un host destino.

Los protocolos de control observados son:

- `ARP`: Funciona en la capa de enlace y permite a los dispositivos mapear direcciones IP a direcciones MAC.
- `LLDP`: Es un protocolo de capa de enlace para la detección de vecinos, permite que los dispositivos anuncien su información a aquellos conectados directamente.
- `1905.1IEEE`: Funciona entre la capa de enlace y la de red, permite a los dispositivos obtener una visión global de la topología de la red, independientemente de las tecnologías con las que esté compuesta.

## 3.2. Servicios de streaming

En este experimento vamos a analizar el tráfico que circula por redes conectadas a internet cuando se utilizan servicios de streaming; algunos de ellos pueden ser YouTube, Twitch, Netflix, etc. Estos servicios son utilizados frecuentemente por los usuarios de internet y representan una gran parte del ancho de banda de una red ya que están constantemente descargando datos de audio e imagen.

El código utilizado para la captura fue el mismo que la sección de modelado. Para este análisis tomaremos muestras de un tamaño de 200.000 tramas. La red utilizada sera de tipo Wi-Fi, en el hogar de uno de los integrantes del trabajo. La captura fue realizada un día Martes a las 18hs, horario en el cual no se hace un gran uso de la red. En las siguientes figuras se pueden observar los resultados obtenidos:

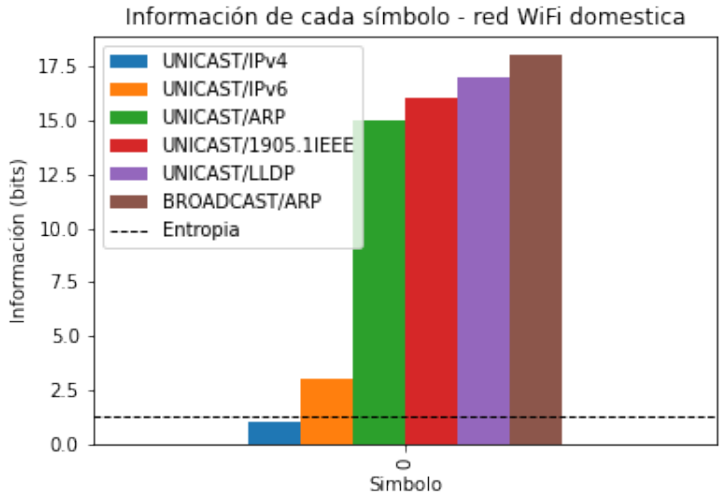


Figura 4: Información de paquetes recibidos utilizando servicios de streaming

```
Tipo de paquete        : Probabilidad de ocurrencia
 'IPv4/UNICAST'        : 0.85835
 'IPv6/UNICAST'        : 0.14156
 'ARP/UNICAST'         : 0.00005
 '1905.1IEEE/UNICAST'  : 0.00003
 'LLDP/UNICAST'        : 0.00001
 'ARP/BROADCAST'       : 0.00001

Entropía de la fuente: 1.28444
```

Observamos que el símbolo distinguido de la fuente es `IPv4/UNICAST` y en menor medida `IPv6/UNICAST`, que tienen asociada una cantidad notablemente menor de información respecto de los demás símbolos de la fuente. Esto esta relacionado directamente con sus altas probabilidades de ocurrencia, como se observa en la tabla. Asimismo vemos que los símbolos con mayor cantidad de información asociada son los de menor probabilidad de ocurrencia.

Creemos que la prácticamente nula aparición de paquetes de tipo `BROADCAST` se debe a que la red es de tipo doméstica, por lo que los dispositivos de la misma se conocen entre sí y las tablas ARP no requieren ser modificadas. En general una red que esta siendo usada recibe paquetes `IPv4/UNICAST` o `IPv6/UNICAST` y la aparición de paquetes de control es menor, por lo que representa en buena medida el comportamiento general de la red.

# 4. Nodos distinguidos

Podemos enfocarnos en un tipo de paquetes mas específico como son los de protocolo `ARP` para intentar distinguir los distintos nodos de una red. Como explicamos anteriormente, ARP es un protocolo de comunicaciones responsable de asociar una dirección a nivel de red (IP) a una dirección física (MAC). Para lograrlo, un dispositivo envía un paquete de tipo `who-has` en formato `BROADCAST` a la red preguntando quien posee la dirección IP que se esta buscando y espera a que el dispositivo que tenga esa dirección IP responda con un mensaje `UNICAST` de tipo `reply`.

En la figura 5 podemos observar el formato de un paquete ARP, algunos de los más fundamentales son:

- **HardwareType:** Especifica el tipo de red física (ej. Ethernet)
- **ProtocolType:** Especifica el tipo de protocolo de la capa de red
- **HLEN y PLEN:** Longitud de las direcciones de Hardware y Protocolo
- **Operation:** Especifica si es un mensaje de solicitud o respuesta

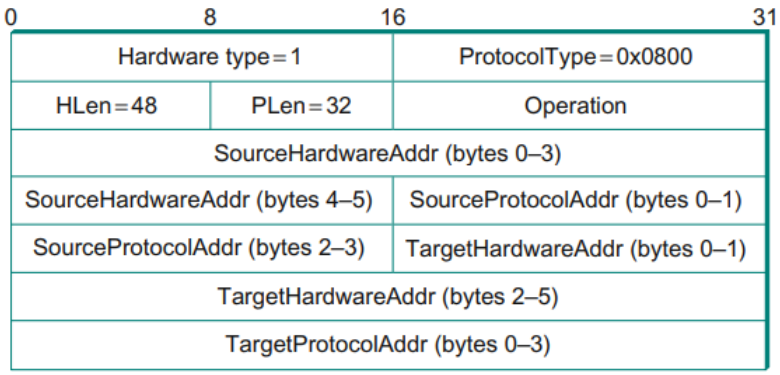


Figura 5: Formato de un paquete ARP

Para realizar esta distinción de nodos dentro de una red planteamos una fuente de información de memoria nula $S2$ a partir de los paquetes del protocolo `ARP` que circulan en la misma. Los símbolos de esta fuente seguirán el formato: `| IP src | Operation |` donde la operación será `who-has` o `is-at`, siguiendo la idea de este protocolo.

A partir de las direcciones IP y el tipo de operación pretendemos obtener información que nos permita identificar los distintos hosts de las redes analizadas. Para la implementación utilizamos la herramienta Wireshark, con la que obtuvimos capturas de 5.000 tramas de tipo `ARP`. Además desarrollamos un script usando Scapy para analizar las capturas a partir de los archivos generados y poder luego graficar los datos obtenidos.

```
# Lectura de paquetes ARP y fuente de información
S2 = {}
for pkt in PcapReader(fileIn):
    arp_op = 'who-has' if (pkt[ARP].op == 1) else 'is-at' # Operation
    s2_i = (pkt[ARP].psrc, pkt[ARP].pdst, arp_op)

    if s2_i not in S2:
        S2[s2_i] = 0.0

    S2[s2_i] += 1.0
```

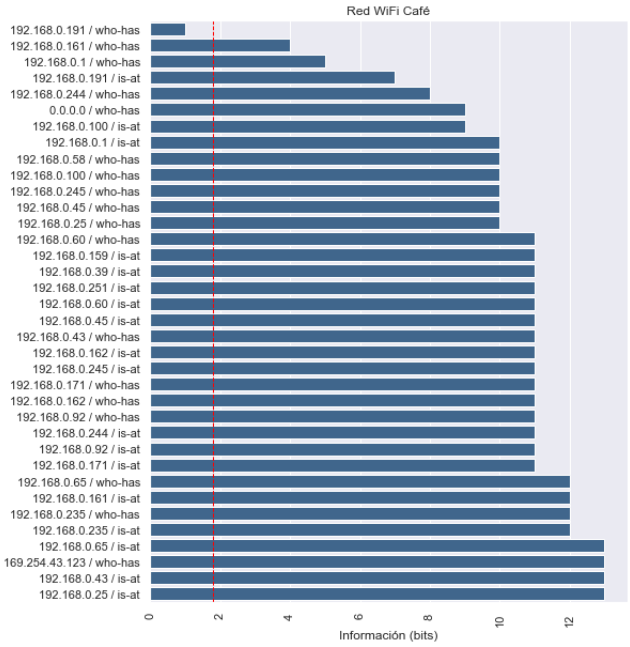


Figura 6: Información de cada símbolo ARP circulante

A partir de las figuras obtenidas podemos observar que nuestra captura obtuvo 36 símbolos distinguidos unívocamente por su IP y tipo de operación. En particular, el símbolo con menor cantidad de información asociada y mayor probabilidad, es `| 192.168.0.191 | who-has |`. Además, vemos que la entropía no es máxima, pero hay conjuntos de símbolos con la misma probabilidad; creemos que esto se debe a que a pesar de que no todos los dispositivos hagan las mismas consultas ARP, cuando un dispositivo envía un paquete ARP con operación `who-has`, el único que va a responder es el que tenga dirección IP igual a la que se esta consultando.

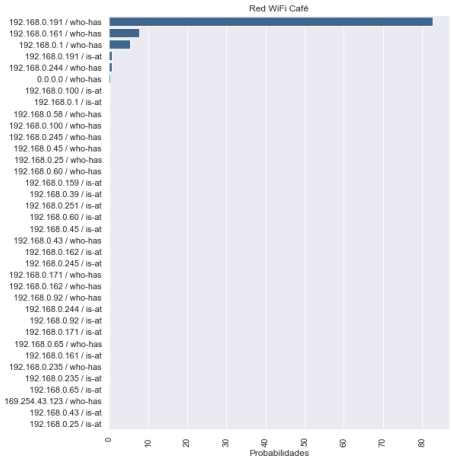


Figura 7: Probabilidad de cada símbolo ARP circulante

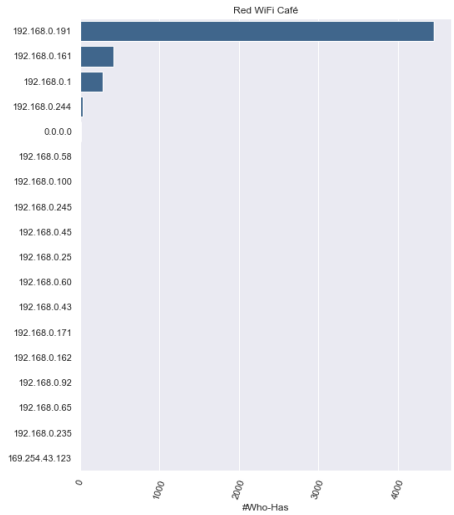


Figura 8: Cantidad de solicitudes de tipo `who-has` por cada IP

Los nodos pueden distinguirse a partir de la dirección IP fuente, ya que esta es única para cada host en la red. Creemos que el host con dirección IP `192.168.0.191` es el router de la red ya que, como se observa en la figura 8, es quien realiza la gran mayoría de consultas sobre la red. El resto de los nodos que generan tanto trafico ARP podemos suponer que son dispositivos como celulares o computadores que están utilizando la red.

En general el comportamiento de la red nos resultó normal, se tiene un router y varios hosts que usan la red de forma independiente. Sin embargo observamos la presencia de un host con dirección IP igual a 0.0.0.0, hemos encontrado que esta dirección se usa como default o como *placeholder*. Por ejemplo en IPv4, la dirección 0.0.0.0 es una meta-dirección no ruteable que se usa para designar un destino invalido o desconocido.

# 5. Conclusiones

A partir de los experimentos realizados creemos haber obtenido un mayor entendimiento acerca del comportamiento de ciertas redes de comunicación entre dispositivos electrónicos, principalmente aquellos que se comunican mediante la Internet.

Vimos que en redes privadas formadas por un conjunto de hosts determinado con poca variación, el tráfico está casi en su totalidad formado por paquetes unicast. Por su parte, redes públicas donde la topología varía constantemente presentan una proporción mucho mayor de paquetes Broadcast. Aún así, en ambos casos la gran mayoría del tráfico está conformado por paquetes de datos y no de control.

Además observamos que, al menos en los casos de estudio analizados, el protocolo `IPv4` continúa siendo más utilizado que su nueva versión `IPv6`.

Relacionando los resultados observados con la teoría de la información de Shannon, no encontramos que se alcanzara la entropía teórica máxima en ninguna de las redes analizadas. Pensamos que una red en que todos los símbolos tengan la misma probabilidad de ocurrencia podría ser una con un propósito específico en que cada tipo de paquete que se envía tiene una respuesta asociada y los mensajes se envían siguiendo una estructura determinada. Este no es el caso de las redes de propósito general que analizamos.