DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico Nro 1

Wiretapping

20 de abril de 2016 — Teoría de las Comunicaciones

Integrante	LU	Correo electrónico
Bouzón, María Belén	128/13	belenbouzon@hotmail.com
Rey, Maximiliano	037/13	rey.maximiliano@gmail.com
Tarantino, Patricio M.	369/10	patriciotarantino@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

Contents

1 Introducción

1.1 Objetivo

El objetivo propuesto por la cátedra fue el de utilizar técnicas provistas por la teoría de la información para distinguir diversos aspectos de una variedad de redes de manera analítica. Para ello, nos fue solicitada la implementación de un conjunto de herramientas que nos posibilitaran capturar, manipular y analizar paquetes de información haciendo hincapié en la observación de los símbolos distinguidos que pudiésemos notar en cada una de las fuentes propuestas y en la entropía de cada una de ellas.

1.2 Marco teórico

Dada la consigna propuesta por la cátedra, existen dos conceptos centrales que resulta pertinente comprender antes de continuar con la lectura del presente informe. Estos son, el de Entropía y el de protocolos ARP.

En Teoría de la información se conoce a la entropía como el promedio de información contenida en cada mensaje enviado por una fuente. Como la cantidad de información está determinada en función de la probabilidad de aparición de cada símbolo posible (siendo menor la información aportada cuando el símbolo en cuestión no es distinguido) se desprende lógicamente que la entropía será inversamente proporcional a la predictibilidad de la fuente. Es decir, que cuanto más predecible sea la fuente, menor será su entropía (y viceversa). Dicho de otra manera, cuanto menos probable sea un evento, mayor información proporcionará su aparición. La fórmula para calcular la entropía de una fuente con n eventos probables es:

$$H(S) = \sum_{i=1}^{n} p_i * (-\log p_i)$$

Por otro lado, las siglas ARP hacen referencia a *Address Resolution Protocol*. Como su nombre lo indica, se trata de un protocolo de la capa *data link* responsable de resolver la dirección de hardware (MAC) que corresponde a una determinada dirección IP, conocida por el emisor.

Para que esto sea posible, es necesario que se puedan identificar dos tipos de paquetes ARP:

- Los paquetes *who-is*, que son distribuidos a todos los equipos de la red mediante broadcast con el fin de poder solicitar su MAC al nodo que posea la IP indicada en el paquete.
- Los paquetes *is-at* transmiten mediante unicast la dirección MAC solicitada - entre otras cosas - al equipo que inició la comunicación solicitándola.

Todos los paquetes ARP contienen diversos campos que especifican:

- Tipo de hardware
- Tipo de protocolo
- Longitud dirección de hardware
- Longitud dirección de protocolo
- Código de operación
- Dirección hardware del emisor
- Dirección IP del emisor
- Dirección hardware del receptor
- Dirección IP del receptor

En este trabajo de investigación nos enfocaremos principalmente en los últimos cinco.

2 Detalles de implementación

Con el fin de llevar a cabo el desarrollo del presente trabajo práctico desarrollamos una serie de algoritmos que hacen uso de las bibliotecas Scapy, un manipulador de paquetes interactivo escrito en Python. Uno de ellos, *escuchar2.py* se encarga de escuchar la red en modo promiscuo y generar un archivo *.pcap con la información de los paquetes transmitidos. Luego, *cuantificacionTiposDePaquetes.py* y *procesarArp.py* toman como input - entre otros parámetros - dichos archivos y otorgan información de dos tipos: el primero informa sobre la cantidad de paquetes observados de cada tipo posible, la información que cada tipo provee y la entropía de la fuente. El segundo, filtra únicamente los paquetes de tipo ARP y calcula cuántas veces cada IP fue requerida en un paquete *who is*, en cuántas ocasiones cada IP envió mensajes de dicho tipo y cuál es la entropía de cada una de las fuentes.
Por otro lado, en el archivo *utils.py* se implementó un mapeo que hace explícita la relación entre un tipo y su representación en valor hexadecimal.

Una serie de gráficos fue generada a partir de los datos obtenidos con el fin de presentarlos de una manera más transparente y de mejor lectura. En ellos se puede distinguir rápidamente la entropía de cada fuente, los distintos valores observados y los nodos que pueden considerarse significativos en cuanto a valor por encima de la entropía (aquellos cuya ocurrencia no es frecuente) o por debajo de la misma (aquellos que aportan poca información debido a su alta frecuencia de aparición).
Otro conjunto de gráficos mostrará, para cada fuente, un grafo asociado en donde se verán los vínculos establecidos entre los diversos hosts. Esta forma alternativa de presentar los resultados permitirá exponer un panorama amplio de la red observada que permita una comprensión más intuitiva del funcionamiento de la misma.

Todos los resultados obtenidos y los gráficos realizados servirán a los fines de generar el análisis que se extiende en la próxima sección.

3 Análisis

En esta sección presentaremos el análisis de los distintos casos de estudio abordados. Cada subsección corresponderá a uno de ellos, abordándolos de menor a mayor tamaño.

3.1 Red doméstica

3.1.1 Nodos intervinientes

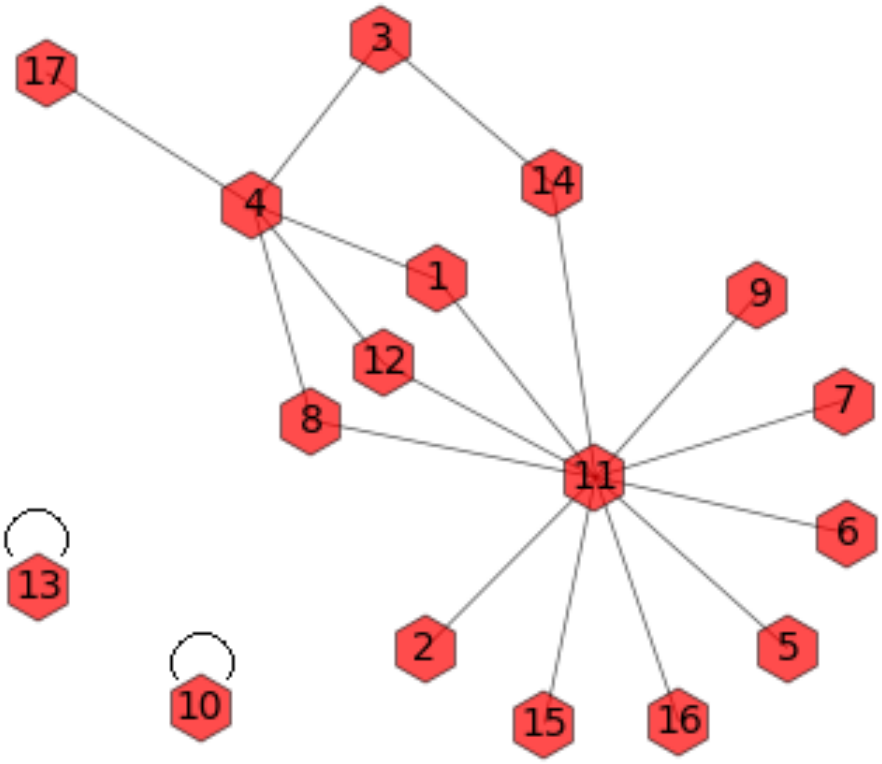

Figure 1: Grafo de Red Doméstica

COMPLETAR

3.1.2 Frecuencia de los distintos tipos de paquetes

En el gráfico 2 se pueden ver los resultados de obtenidos.

3.1.3 Información de las fuentes emisoras y receptoras de paquetes ARP.

Para realizar este caso comparamos la información proporcionada por las fuentes emisoras de paquetes ARP e ilustramos asimismo la entropía de la fuente. Los resultados se pueden observar en los gráficos 3 y 4

En el primero de ellos se ve que existe una única fuente situada por debajo de la entropía y corresponde a la IP 169.254.93.30. La baja información que aporta indica que se trata de un nodo que emite muchos paquetes ARP. En contraposición, las IP 192.168.0.103 y 192.168.0.100 muestran un gran aporte de información, de lo cual se deduce que es poco común que las mismas envíen paquetes de solicitud ARP.

En el segundo gráfico se observa una entropía mayor. De esto podría deducirse que es una fuente aún más aleatoria y menos predecible que la que filtra únicamente los envíos de paquetes ARP.

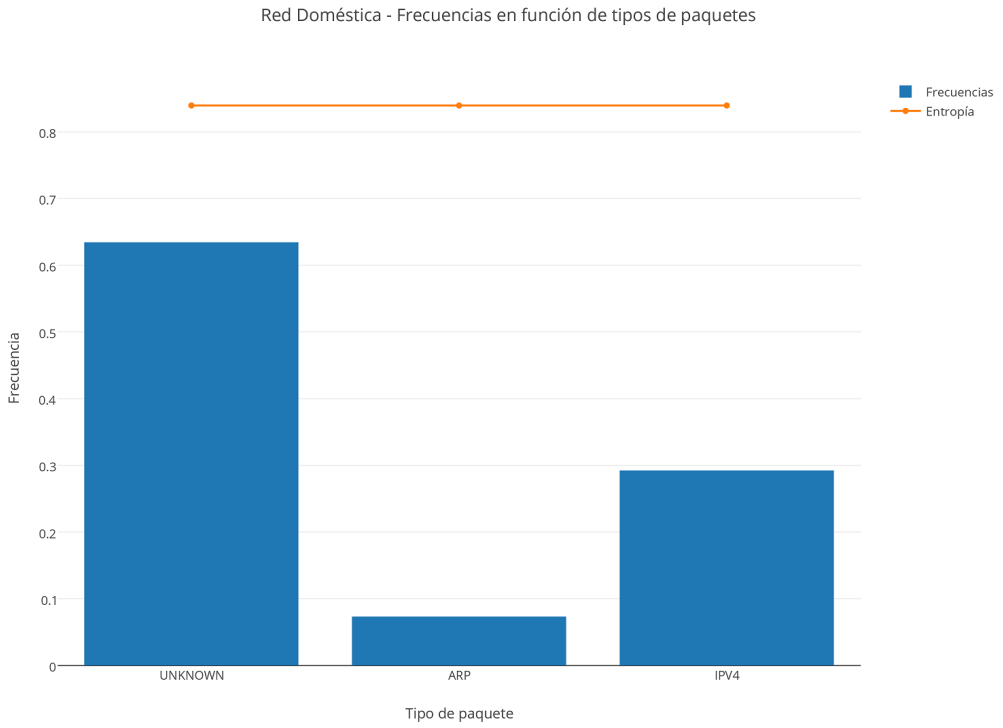

Figure 2:

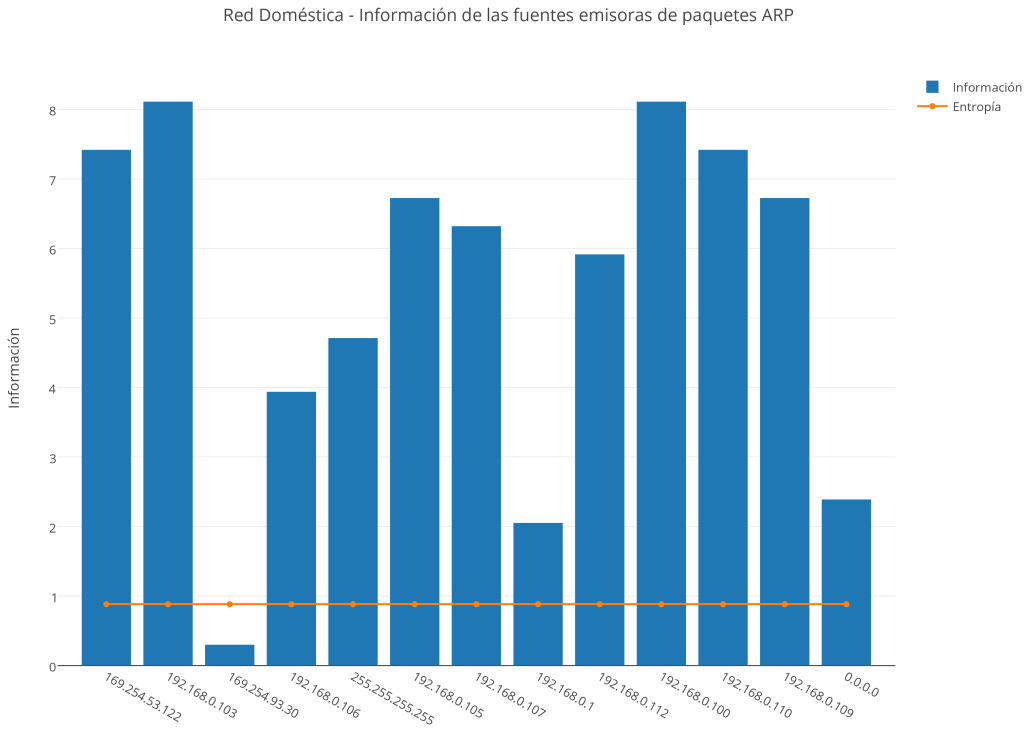

Figure 3:

En este caso, los nodos distinguidos son - por su gran valor de información - los que mapean las direcciones IP 192.168.0.103, 192.168.0.9 y 192.168.0.111 y, por su poca cantidad de información el nodo 0.0.0.0, seguido del 169.254.93.30. Esto indica que frecuentemente se realizan requests a estos últimos

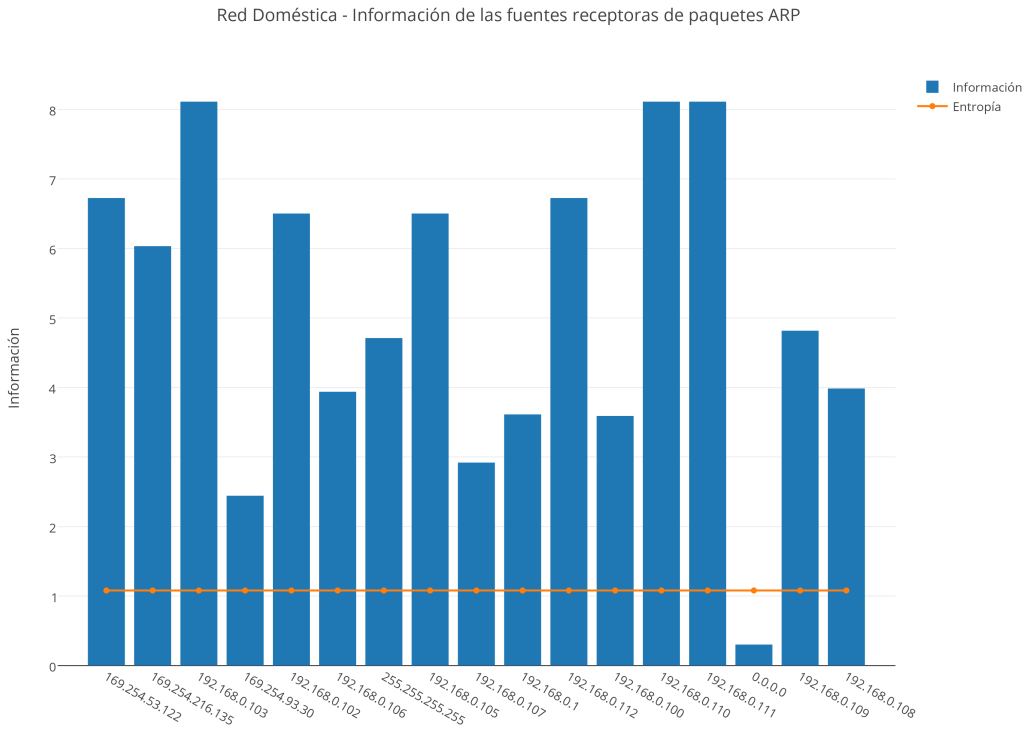

Figure 4:

hosts, mientras que los primeros son ignorados la mayor parte del tiempo.

Por qué es tan frecuente que se realicen solicitudes al host cuya IP es 0.0.0.0? Podríamos pensar que se trata de comunicaciones que un nodo envía a sí mismo pero también cabría suponer - de acuerdo a lo investigado - que puede tratarse de un mecanismo de chequeo de IP's duplicadas en la red.

Comparando ambos gráficos podríamos decir que es bastante común que la IP 169.254.93.30 envíe y reciba paquetes ARP. Es por eso que consideramos que podría ser un potencial candidato a Router.

Viendo otros datos calculados, observamos que el muestreo de la red detectó un alto índice de pedidos de dirección por parte del dispositivo con IP 192.168.0.1, el cual contactó a varios dispositivos en múltiples ocasiones. Puntualmente, se destacó el dispositivo con la IP 192.168.0.107, en el cuál hubo 180 solicitudes, contra una 6 solicitudes por parte del 192.168.0.107 al 192.168.0.1.

Por otro lado los dispositivos con la IP 192.168.0.106 y 192.168.0.108 enviaron una alta cantidad de solicitudes al 192.168.0.1 (respecto de la media).

Viendo esto, podemos decir que el dispositivo con IP 192.168.0.1 es un nodo importante en la red, por lo cual pensamos que puede ser otro candidato a ser el Router. Se puede deducir que durante un periodo prolongado el dispositivo 192.168.0.107 estuvo extrayendo información del nodo, lo cual explicaría por qué existe un trafico de paquetes más elevado desde el nodo hasta el 192.168.0.107 que desde este último al nodo..

En el caso del dispositivo con IP 192.168.0.106, y 192.168.0.108 las solicitudes para buscar al nodo podrían deberse a que estos estaban enviando información al nodo.

La mayoría de los dispositivos contenía una IP con el prefijo de 192.168.0, con lo cual se puede deducir que ésta es la IP de la red analizada. Además, se han detectado paquetes provenientes de redes privadas distintas (puntualmente de tres dispositivos).

Se registró un trafico mínimo entre dispositivos en donde el 192.168.0.1 no estuvo involucrado.

En la red analizada existían dos fvble al router wifi y el resto estaban enlazados al mismo utilizando el wifi, incluso la computadora en la cual se ejecutó el algoritmo de captura de paquetes. Esto podría explicar la presencia de dos redes en las mediciones, una de ellas (169.254) es la red formada por las dos computadoras enlazadas por cable y el proxy que se comunica con internet y la otra red (192.168.0) es la conformada por los dispositivos WiFi. Si esto es cierto el 169.254.93.30 podría ser el proxy y el

192.168.0.1 el nodo que rutea a los dispositivos de la red de wifi a la red del proxy.
El grafo muestra que el IP 0.0.0.0 (nodo 4) tuvo contacto con cinco nodos. Puntualmente El 0.0.0.0 envió datos a cuatro nodos cuyos IP tuvieron actividad con otros nodos de la red, con lo cual se puede deducir esos mensajes provenian de los dispositivos reporesentados por esos cuatro nodos, que intentaban determinar si existían duplicados de su IP en la red.
El nodo 169.254.93.30 es el único envía paquetes al 0.0.0.0, lo cual puede tratarse de un mecanismo de IP para enviar información a todas las redes.

3.2 Red Laboral

En este segundo caso, se analizó una red laboral de una PyME. La captura se realizó por un lapso de media hora en modo promiscuo.

3.2.1 Nodos intervinientes

En el gráfico 5 se pueden ver los resultados de obtenidos.

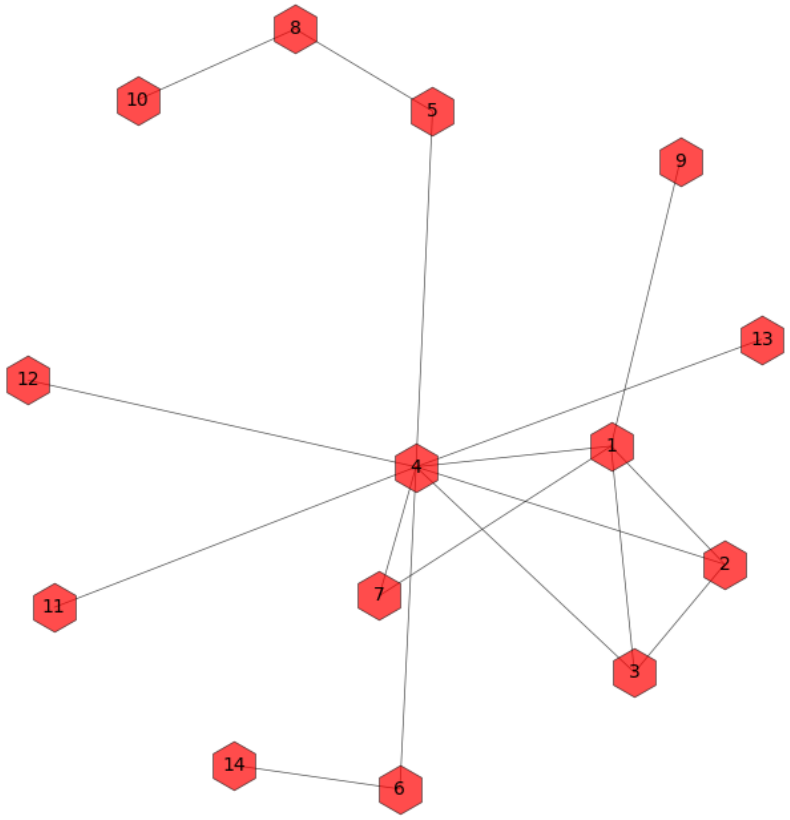

Figure 5: Grafo de Red Laboral

La visualización del grafo indicaría, a primera vista, que el nodo 4 se transmite paquetes con la mayoría de los demás vértices (excepto casos particulares) siendo el candidato ideal para ser el router de esta conexión - es decir, el nodo con salida exterior.

El nodo 1 también se comunica con varios nodos a la vez - esto se debe a que es un servidor local en la oficina, al cual le llegan peticiones de nodos particulares - que a su vez también se comunican con los otros que acceden a dicho servidor local. Es decir, el nodo 2, 3, 7 y 9 acceden al servidor local 4, pero a su vez entre el 2 y el 3 se comunican entre ellos.

Los nodos 10, 8 y 14 parecieran ser nodos particulares que no se conectan con el router. En particular, el nodo 8 es 0.0.0.0, por lo que en realidad es el nodo 5 y 10 comunicándose con ellos mismos . El nodo 14 es una IP interna así que posiblemente es un nodo comunicándose con una interfaz propia.

Luego de ese análisis, consultamos las IPs con la configuración de la red (a la que tenemos acceso por ser la red de la oficina laboral). Efectivamente, el nodo 4 es el router, así como el nodo 1 el servidor local estimado, y los nodos 2, 3, 7 y 9 los que se comunican con dicho servidor.

La incertidumbre del nodo 14 se resolvió como una Virtual Network Adapter, producto de una Virtual Box corriendo en el nodo 6.

Todas las demás relaciones parecen ser naturales, de nodos que se comunican con el router para poder salir a internet.

3.2.2 Frecuencia de los distintos tipos de paquetes

En el gráfico 6 se pueden ver los resultados de obtenidos. Allí se observa la predominancia de los paquetes de tipo IPV4. Esto tiene sentido, puesto que en este contexto es más frecuente el intercambio de paquetes a nivel de Internet que el realizado a nivel de la intranet, con la cual las computadoras se encuentran conectadas a través de Ethernet. Esto también considerando que las mediciones fueron realizadas en un momento de mucha actividad de pasaje de paquetes en la red.

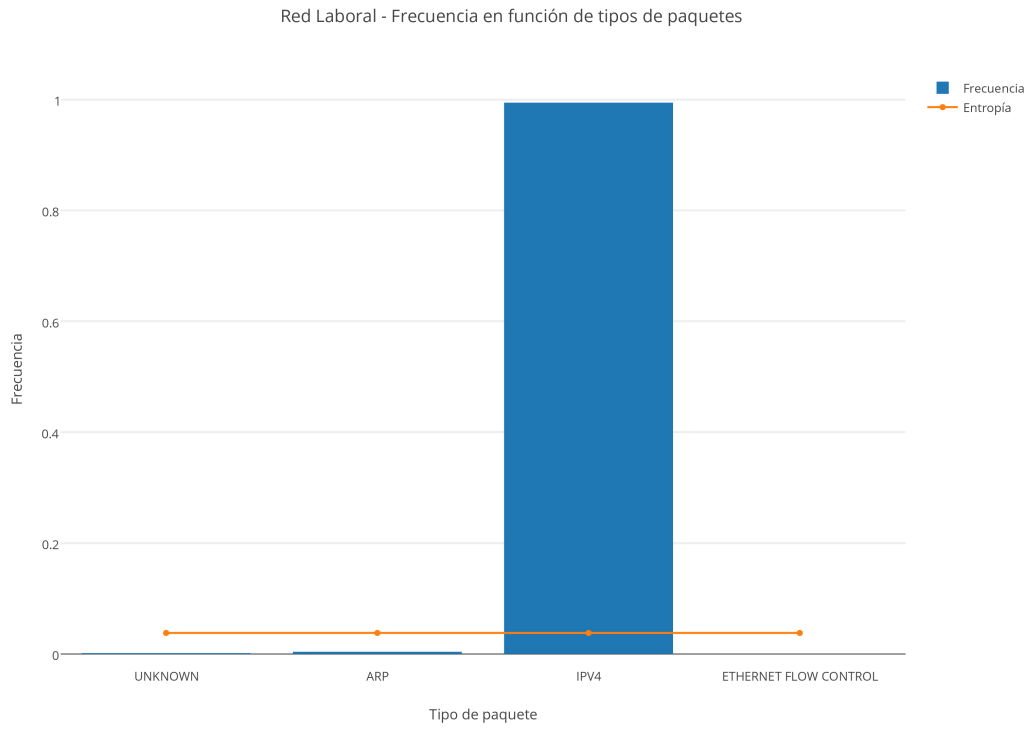

Figure 6:

Comparando las figuras 7 y 8 se puede ver que las fuentes receptoras superan en cantidad a las fuentes emisoras.

3.2.3 Información de las fuentes emisoras y receptoras de paquetes ARP.

Dentro de las IPs emisoras se puede destacar la participación activa de el nodo 192.168.1.1, la cuál aporta un valor muy pequeño de información, encontrándose por debajo de la entropía. Esto indicaría que se trata del host que más paquetes ARP envía dentro de la red - y esto es correcto, ya que en efecto es el nodo 4, el nodo correspondiente al router.

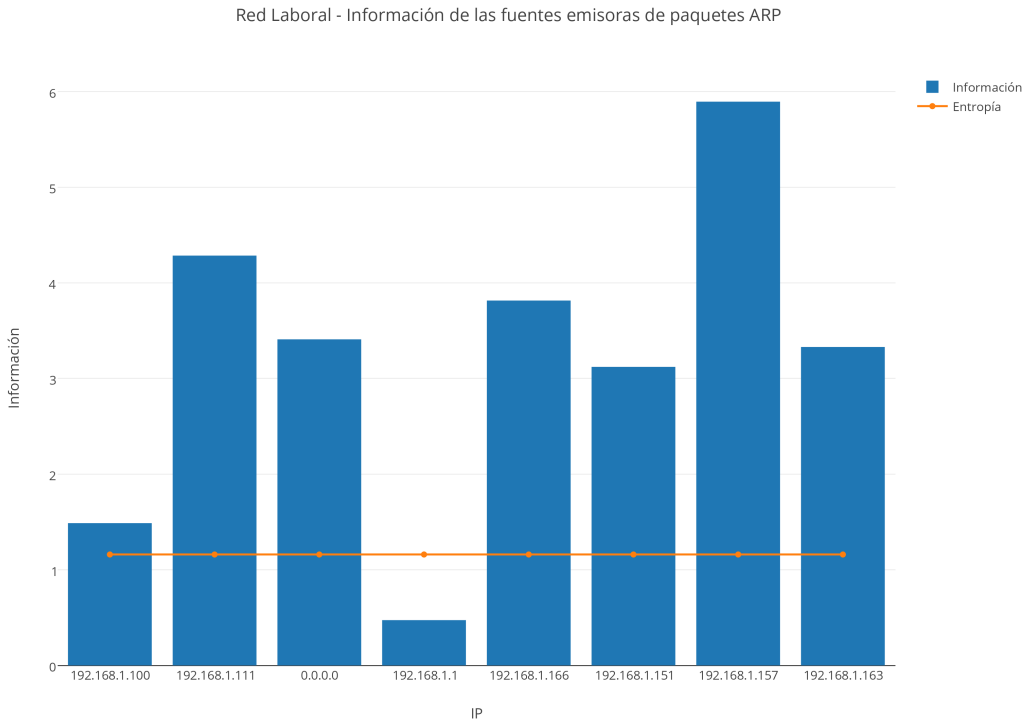

Figure 7:

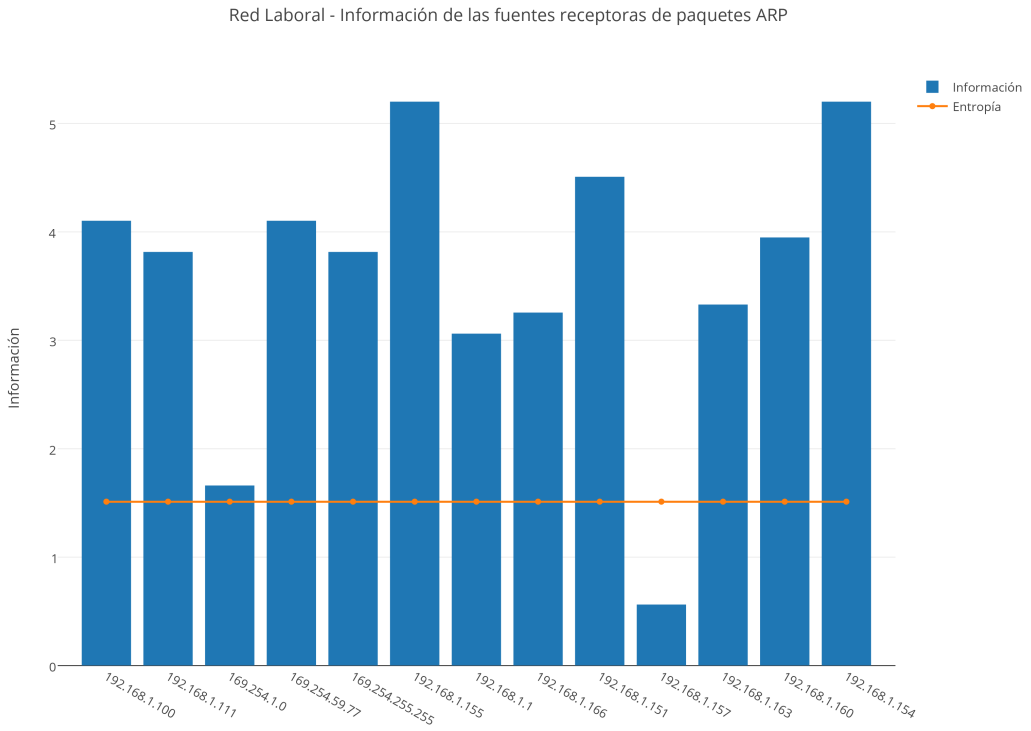

Figure 8:

Si lo buscamos en el gráfico 8, el mismo nodo aporta una información por encima de la media, pero sin ser destacada. Es decir, que hay nodos que envían más paquetes que el router. Esto es porque, si bien el router es salida de todos los nodos a internet, las conexiones internas (es decir, entre nodos, con

el servidor local, por ejemplo) tienen más peso, y se usan esas conexiones más que la red externa.

Otro caso a destacar es el de la IP 192.168.1.157 (nodo 7). Comparándolo en ambos gráficos, se lo ve como el punto máximo en uno y el mínimo en otro: el nodo 7 es aquel que más paquetes envía y el que más recibe, conectándose tanto con el router como con el servidor local. Si bien no pudimos verificar la identidad de dicha IP (pues utilizamos NAT para ciertas IPs y esta cae en una de ellas, siendo imposible identificar cuál era su identidad al momento de la medición), parece sensato suponer que se debe a un nodo (computadora) trabajando con el servidor y el router a la vez, pudiendo deberse la gran cantidad de paquetes enviado por estar subiendo/copiando material a uno de ellos.

3.3 Red Bonafide

Por último, dadas las pequeñas dimensiones de las redes anteriores, decidimos tomar datos de la red pública ofrecida por el local Bonafide. La captura se realizó por un lapso de media hora en modo promiscuo.

3.3.1 Nodos intervinientes

En el gráfico 9 se pueden ver los resultados de obtenidos.

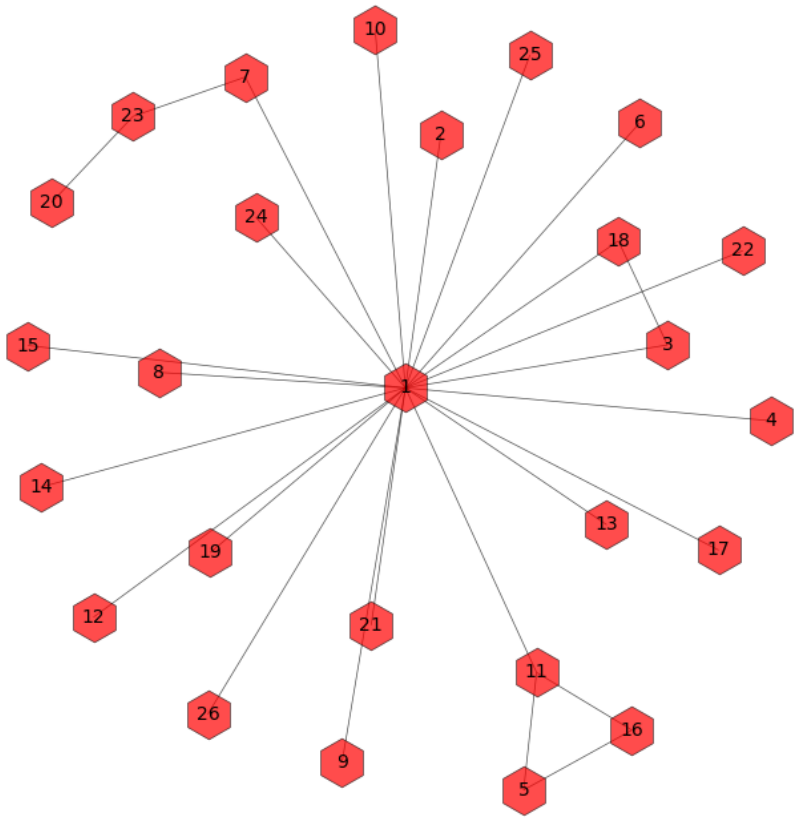

Figure 9: Grafo de Red Bonafide

Si bien el grafo resultante no es exactamente la idea previa que teníamos de él (un grafo estrella con el nodo root como raíz) resulta interesante observar la indiscutible predominancia de aristas incidentes al nodo 1, descripto con la dirección IP 192.168.1.1: el resto de los nodos son adyacentes a él o tienen un camino al mismo de - a lo sumo - tercer grado.
Consideramos que estos datos son suficientes como para deducir que dicha dirección IP coincide con la del router.
Al indagar acerca de la naturaleza de los únicos nodos que no se encuentran conectados de forma directa al nodo 1 (el 5 (192.168.1.110), el 16 (192.168.1.4), el 23(0.0.0.0) y el 20 (169.254.201.232)) llegamos a las siguientes conclusiones:

- Tanto el nodo 5 como el 16 (ambos de clase C) se comunican con el router a través del nodo 11 (192.168.1.105). Ahondando en las características de estas transferencias, se puede ver que el nodo

11 realiza requests a los nodos 1, 5 y 16, pero sólo llega a él un request desde el nodo 16. Podríamos entonces inferir que el nodo 11 puede estar funcionando como switch.REVISAR

- El nodo 20 se conecta al 23 y éste último al 7 (192.168.1.100). Esto forma la secuencia 169.254.201.232 (clase B)- 0.0.0.0 - 192.168.1.100 (clase C) - router. Si bien no se trata de un grafo dirigido, se puede ver en los resultados que la dirección 0.0.0.0 nunca recibe datos que podamos escuchar, si no que envía requests a ambos nodos que lo circundan.
  De acuerdo a lo investigado, inferimos que se trata de un fenómeno que ocurre cuando un nuevo host se conecta a la IP, con el fin de verificar que no tenga una dirección que se encuentre en uso, evitando así las IP duplicadas.

3.3.2 Frecuencia de los distintos tipos de paquetes

En el gráfico 10 se pueden ver los resultados de obtenidos.

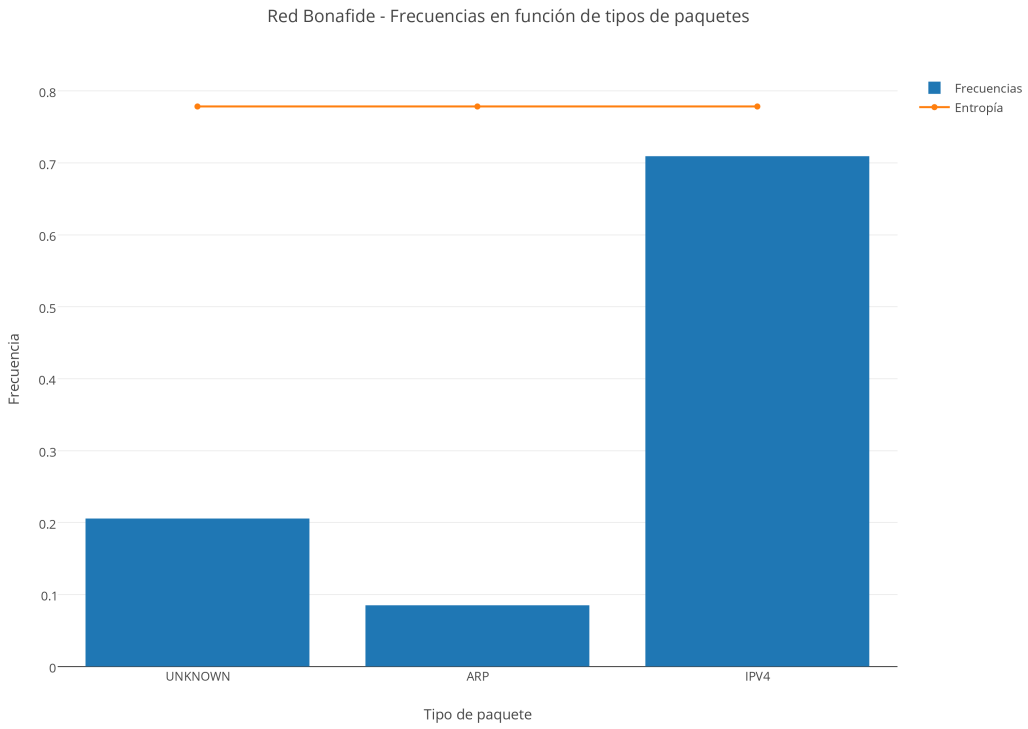

Figure 10:

En este caso se ve una amplia predominancia de los paquetes de tipo IPv4. Esto se corresponde con nuestras expectativas, puesto que los paquetes de tipo ARP se envían como forma de reconocimiento entre dispositivos. Una vez que estos reconocimientos fueron establecidos, el caudal de paquetes de transferencia IPv4 aumenta, superando así a los paquetes ARP en una red quya estructura no se modifica a cada momento debido a la conexión y desconexión de dispositivos.

3.3.3 Información de las fuentes emisoras y receptoras de paquetes ARP.

Este caso podría ser el de más arduo análisis, puesto que es el que mayor caudal de datos presenta. Sin embargo, la gran cantidad de información que contiene facilita la distinción de los nodos significativos.
‘ Como se puede observar en la figura 11, hay cuatro nodos que merecen nuestro reconocimiento: el 7 (192.168.1.100), el 10 (192.168.1.104), el 1 (192.168.1.1) y el 22 (192.168.1.129). Estos aportan un nivel de información menor al valor de la entropía, indicando, en este caso, que se encuentran enviando una gran cantidad de parquetes ARP. Esta deducción no se podría haber hecho analizando únicamente el grafo que representa la red, dado que allí tres de dichos nodos no parecen destacarse de los demás.

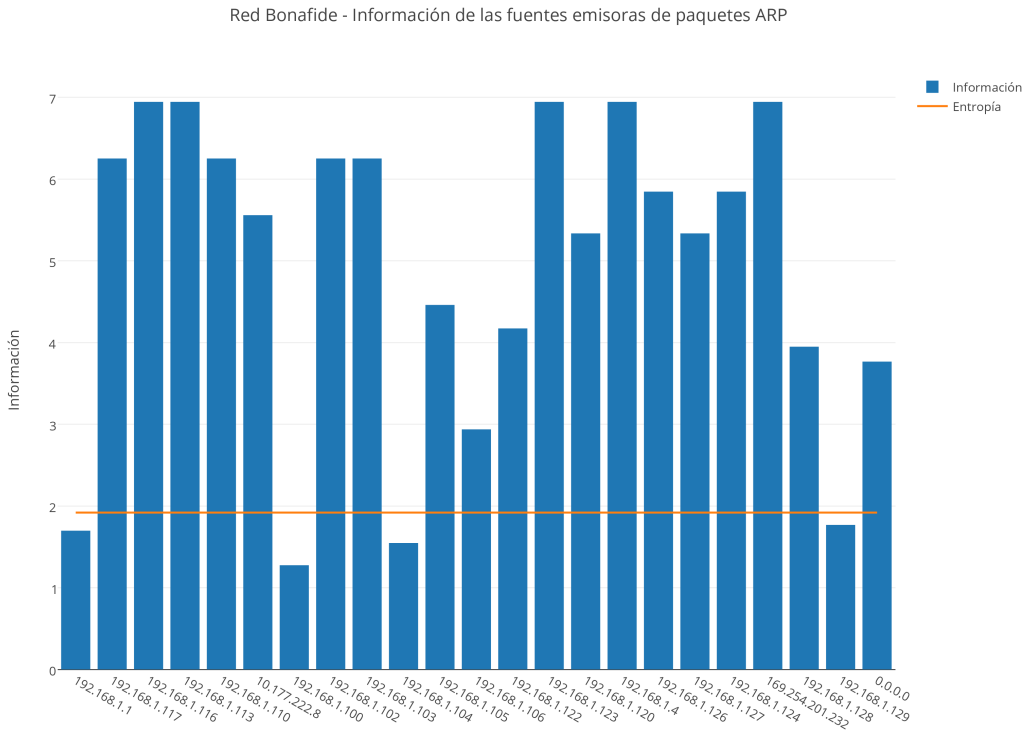

Figure 11:

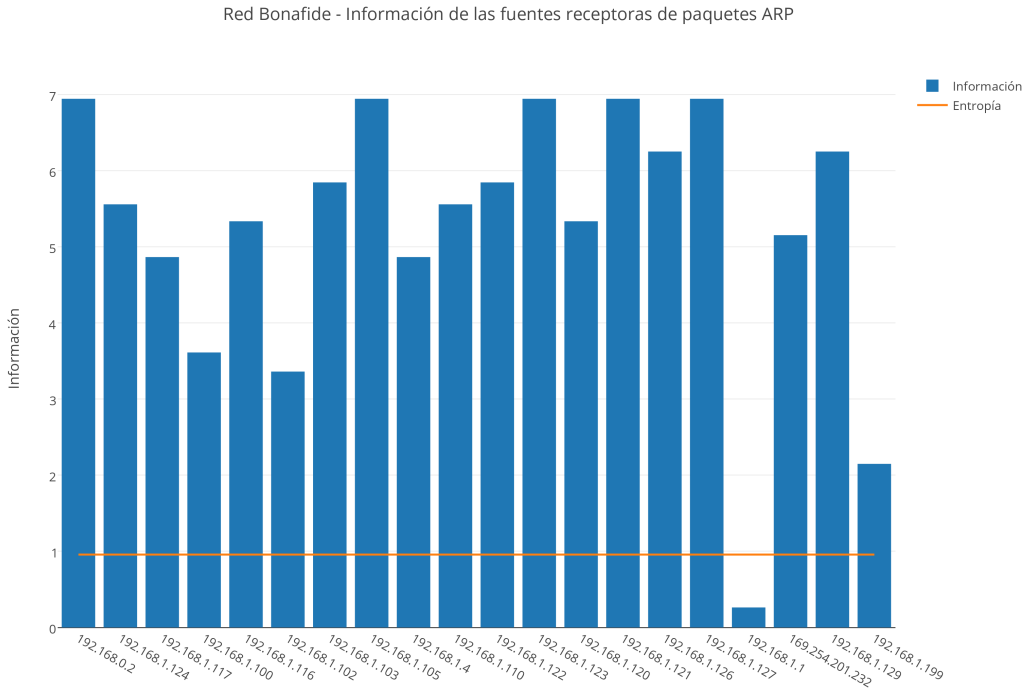

Figure 12:

En cambio, en la figura 12, se refuerza fuertemente lo que se percibe en el grafo de la imagen 9: La mínima información está vinculada a la IP 192.168.1.1, es decir, al nodo identificado con el número 1. Con los tres gráficos presentados en conjunto es posible concluir que la IP 192.168.1.1 es la que identifica al router, no sólo por la cantidad de paquetes ARP que envía y recibe sino también por la estructura de

la red que se conforma en torno al dispositivo.

4 Conclusiones

Al comenzar a resolver los problemas presentados teníamos una idea poco pulida acerca de las redes en general. Conjeturábamos ideas respecto a sus estructuras y comportamientos que se vieron refutadas durante el desarrollo y análisis del presente trabajo práctico. Por poner un ejemplo, esperábamos encontrar redes cuyo grafo fuese un árbol estrella, con el centro como Router y donde no hubiese ciclos ni conexión alguna entre los nodos secundarios. Asimismo, hicieron su aparición direcciones IP que no teníamos presentes.
Luego de concretar los experimentos realizados presentamos sus resultados de una forma que consideramos apropiada estudiamos sus resultados, destacando nodos que consideramos distinguidos, investigando el motivo de los resultados imprevistos, obteniendo datos (como la información de diversos eventos y la entropía de una variada cantidad de fuentes) y reuniendo para el análisis integral los datos proporcionados por distintas herramientas que generamos.
Luego de todo este proceso, concluimos que:

- el cálculo de la entropía resulta imprescindible para conocer la predictibilidad de la fuente que se estudia.
- Es posible deducir con cierta confianza cuál es el Router de una red a partir de las interacciones en las que se involucra pasiva o activamente cada host: es, generalmente, el que mayor actividad tiene por su alta intervención como mediador entre máquinas.
- El protocolo ARP es fundamental para la traducción entre direcciones de nivel de red (IP) y direcciones de nivel de enlace (MAC).