DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico #1

24 de Septiembre de 2013

Teoría de las Comunicaciones

Integrante	LU	Correo electrónico
Nahuel Lascano	476/11	laski.nahuel@gmail.com
Agustin Martinez Suñé	630/11	agusmartinez.92@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

1. Introducción

El objetivo del presente trabajo es el análisis de un protocolo de envío de paquetes en una red. En particular analizaremos el protocolo ARP, primero capturando este tipo de paquetes en diferentes redes y luego analizando esos datos para sacar conclusiones acerca del comportamiento de esas redes.

Además, como objetivo secundario, se nos planteó que usemos el tráfico de paquetes ARP de la red como una fuente de información para poder estudiarla con las herramientas vistas en la materia sobre teoría de la información. Para esto debimos definir un conjunto de símbolos y utilizar los datos de las capturas para calcular la entropía de las redes. Esto permitirá caracterizar con más herramientas a los nodos de la red.

2. Capturando tráfico

Para poder entender el trabajo que hicimos es preciso en primer lugar comprender el funcionamiento del protocolo ARP. El protocolo se utiliza para averiguar a qué dirección MAC corresponde una dirección IP determinada dentro de una red. El nodo que quiere averiguar esta información hace un broadcast Ethernet de un paquete ARP "who-has" con la dirección IP a consultar, y alguno de los nodos de la red que tenga la respuesta en su tabla ARP le envía un paquete ARP "is-at" con la dirección MAC correspondiente.

La herramienta que desarrollamos captura todos los paquetes de tipo ARP que circulen en la red mediante el comando $sniff()$ del paquete Scapy. Por cada paquete guardamos diferentes datos: la IP destino (de dispositivo nodo se quiere averiguar la MAC), la IP origen (qué dispositivo es el que quiere averiguarla), y el tipo de paquete (who-has o is-at). De esta manera, la fuente de información que queda definida es la red visible al nodo desde el que estamos capturando paquetes, y el conjunto de simbolos de esta fuente son las direcciones IP de los nodos.

Mediante la captura de los paquetes vamos a poder caracterizar cada nodo de la red, sabiendo qué nodos se conectan más frecuentemente con qué otros nodos, y además vamos a poder calcular cuál es la entropía de la red.

Realizamos la captura de paquetes de dos fuentes diferentes. En primer lugar recolectamos durante una hora los datos de los paquetes de la red Wi-Fi de los laboratorios de la facultad, y por otro lado recolectamos datos de los paquetes de una red montada en el lugar de trabajo de un compañero.

3. Análisis de las capturas

Para el análisis de los resultados de la recolección de datos utilizamos diversos métodos. En primer lugar, tomamos la red como una fuente de información y las IPs como símbolo. Usando esto como modelo, calculamos la entropía de las redes invesigadas, usando tanto las ips de los receptores como las de los emisores de los paquetes ARP. Otro modelo de información posible, que también nos permitiría caracterizar los nodos de la red y su frecuencia relativa de aparición, sería utilizar como símbolos las direcciones MAC de los paquetes. Esto nos da otro tipo de información, porque muchas veces pueden ser distintos

Luego, para tener una mejor visualización de los datos obtenidos, decidimos utilizar diferentes gráficos para sintetizar y presentar la información que consideramos más relevante de las redes analizadas.

En primer lugar usamos gráficos de torta para representar la proporción de apariciones entre las IPs, tanto de emisores como receptores. Sin embargo, cuando el número de IPs distintas crecía (en particular con las redes de la facultad) este método dejaba de ser eficiente porque el gráfico se volvía incomprensible. Es por esto que en ambas capturas mostramos una versión "reducida" del gráfico, con los primeros 50 paquetes en lugar de todos los capturados.

En segundo lugar, usamos gráficos de barras para analizar la frecuencia absoluta y relativa de aparición de cada una de las IPs, contando por separado los paquetes en los que la IP era emisora y receptora.

Por último, usamos un histograma en 2 dimensiones con una escala cromática que indica la cantidad de

veces que la IP de la fila x envió un ARP preguntando por la IP de la columna y.

Para ahorrar espacio con el texto de los gráficos, en todos los casos salvo los gráficos de torta las IPs se expresan incompletas (solo los últimos 8 bits, que son los que cambian). La primera parte en el caso de los laboratorios de la facultad es siempre "10.2.100" y en la del trabajo de nuestro compañero "10.10.99".

Los resultados y análisis se exponen a continuación.

3.1. Laboratorios del DC

Con un total de 200 paquetes ARP capturados, obtuvimos una entropía en los emisores de 2.81763365265 y en los receptores de 4.20465276111.

3.1.1. Análisis de entropía

Una primera observación que podemos hacer es que la entropía de los receptores es significativamente mayor que la de los emisores. Esto quiere decir que hay más incertidumbre en quién va a ser receptor que en quién va a ser emisor, esto probablemente se deba a que hay menos nodos que funcionan como emisores en comparación con los que funcionan como receptores, o que hay un conjunto reducido de nodos que tienen una alta probabilidad de ser emisores. Las figuras 1 y 2 nos permiten aclarar un poco más estas hipótesis. Vemos que más de la mitad de los paquetes ARP tiene como origen a la IP 254, y el resto del espectro (un poco menos de la mitad) está dividido en las IPs restantes con una pequeña preponderancia de algunas. En cambio la distribución de las IP destino es totalmente fragmentada. Si bien la IP 254 sigue siendo la mayoritaria no tiene la mayoría del espectro, que se acerca a una división más equitativa que las IPs origen. El hecho de que la distribución de IPs origen sea más equitativa hace que el nivel de incertidumbre (y por tanto su entropía) sea mayor, recordemos que la entropía máxima se alcanza con una distribución equiprobable de los simbolos.

Más adelante, junto con la información que nos brindan los otros gráficos, vemos que correlato tiene esté análisis "abstracto" de la entropía con la función que cumplen los nodos en la red.

3.1.2. Análisis de capturas

Como se puede observar claramente en las figuras **??** y 3, hay una gran presencia del dispositivo con la IP 10.2.100.254 (que genera muchas más solicitudes de las que responde), seguido de lejos por la 10.2.100.98.

Esto también se puede apreciar en la figura 4, donde se aprecia que 10.2.100.98 prácticamente solo pregunta por 10.2.100.254. También en el histograma en se puede visualizar que 10.2.100.254 pregunta por varias IPs distintas y también es consultado por varias otras. Otra observación, bastante extraña por cierto, es que hay cierta tendencia de algunos dispositivos a preguntar por ellos mismos (por eso la clara diagonal del histograma).

En las figuras 3 y 4 se observa que la mayor parte de las IPs disponibles (tomando en cuenta solo las que comienzan por "10.2.100") fueron usadas en algún momento de la captura.

En contadas ocasiones (apreciables tanto en el dump de la captura como en las figuras 3 y 4), apareció la IP 0.0.0.0 (la única que no comienza con "10.2.100") como emisora de solicitudes ARP con distintos destinos. Investigando un poco, descubrimos que esta es una IP que se utiliza para dispositivos que no están conectados a ninguna red TCP/IP. Es también una IP ficticia que se usa para representar direcciones inválidas, a menudo impresoras mal configuradas, paquetes con emisor desconocido, o cuando DHCP falla o tarda demasiado tiempo.

También descubrimos que, salvo por las consultas autorreferenciales y las que involucran a 10.2.100.254 y a 0.0.0.0, las consultas entre nodos distintos son inexistentes.

A raíz de estas observaciones, podemos concluir algunas cosas:

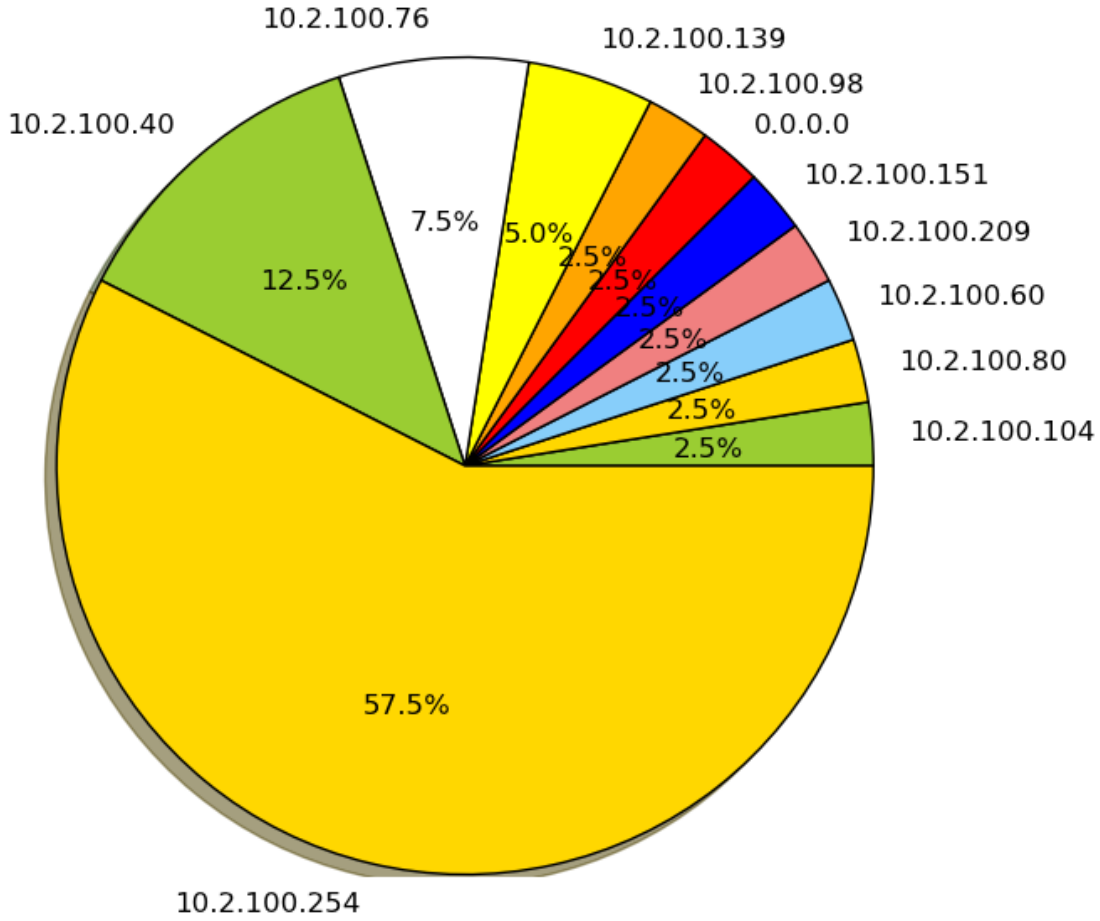

Figura 1: Proporción de IPs origen en los primeros 5 minutos de captura

- El router de la red seguramente se encuentra en 10.2.100.254, y probablemente sea el nodo de salida a Internet.
- En 10.2.100.98 probablemente haya un dispositivo que estaba configurado para renovar sus tablas ARP en intervalos cortos.
- Hay ocasiones en las que DHCP falla y deja dispositivos sin IP, o bien hay dispositivos enviando paquetes con IPs emisoras inválidas o secretos. En cualquier caso, esto arruina la función de ARP, ya que nadie puede responder a una IP que no es ruteable.
- Por momentos los dispositivos tienen leves síntomas de esquizofrenia.
- Más allá de las preguntas por sí mismos y las consultas al router, no existen consultas entre dispositivos. Mientras duró la captura ninguno de ellos trató de comunicarse con otro.
- Todo esto nos lleva a pensar que la función principal de la red es proveer de salida a Internet a los dispositivos que se conecten a ella, más que permitirles la comunicación directa entre dispositivos de la misma red.
- La cantidad de dispositivos conectados a la red es grande. Esto podría explicar la sobrecarga del servidor de DHCP, que falla o tarda mucho tiempo en asignar IPs a los nuevos dispositivos.

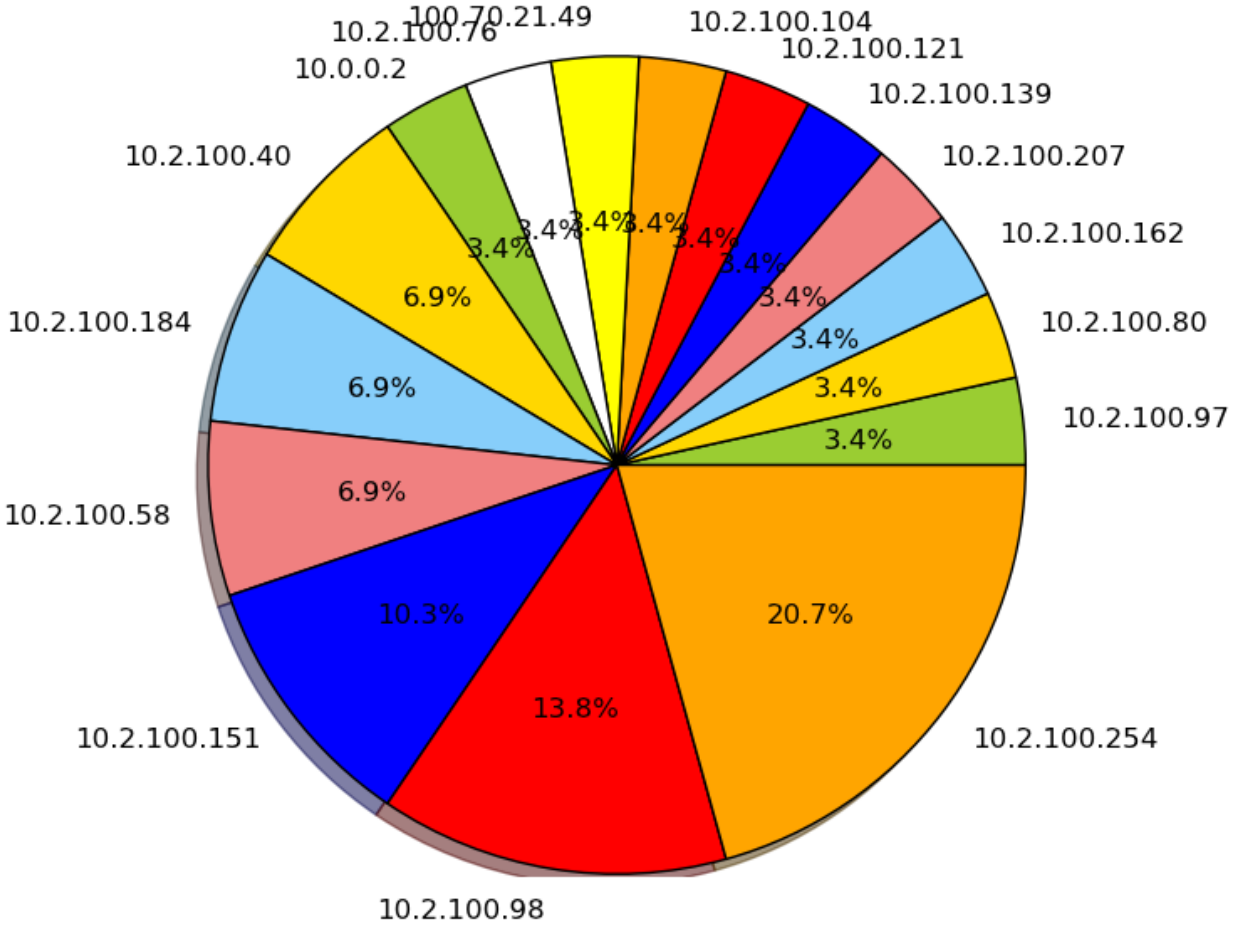

Figura 2: Proporción de IPs destino en los primeros 5 minutos de captura

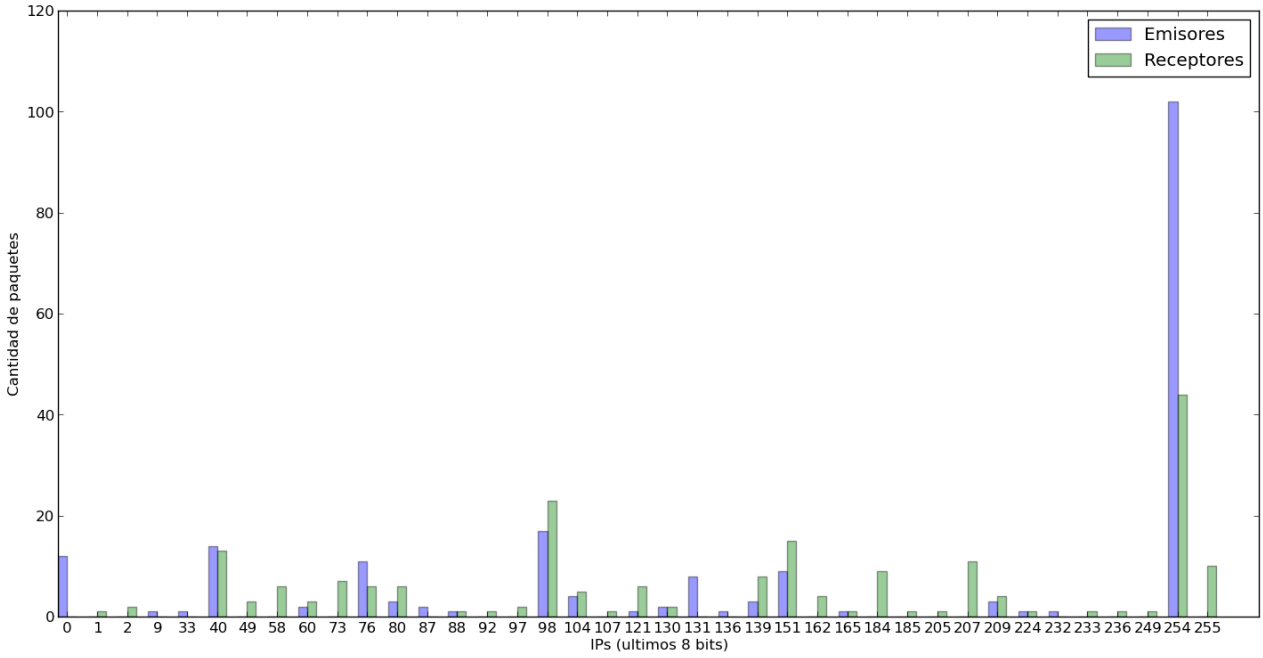

Figura 3: Cantidad de paquetes por IP discriminados en emisores y receptores

3.2. Oficinas de InvGate

Con un total de 11158 paquetes ARP capturados, obtuvimos una entropía en los emisores de 2.85760098357 y en los receptores de 3.08636573523.

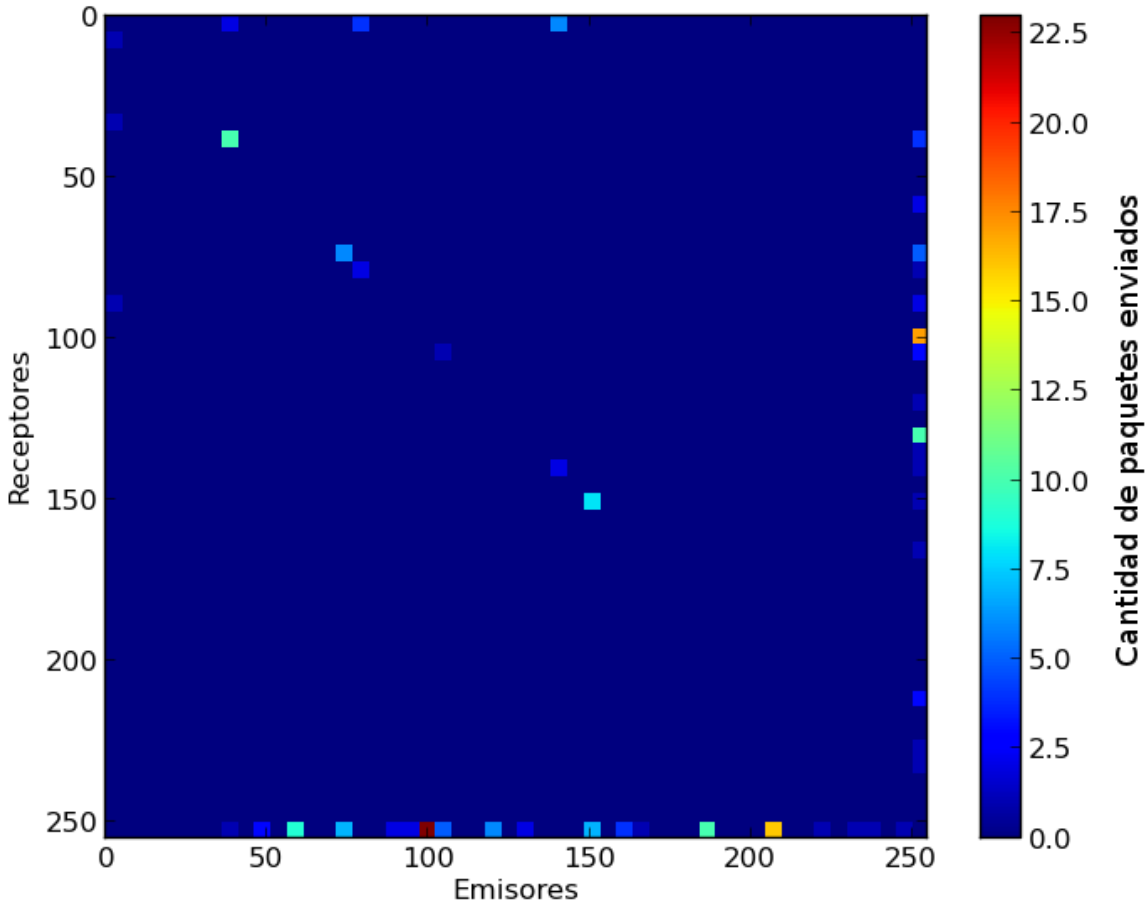

Figura 4: Relación entre emisores y receptores

3.2.1. Análisis de entropía

En este caso no hay una diferencia significativa del valor de la entropía entre IPs origen e IPs destino. Pero así como anteriormente contrastamos estos valores con los gráficos podemos hacer lo propio con este caso. En el gráfico de IPs origen vemos que básicamente la mitad del espectro está acaparada por una computadora y la otra mitad está distribuida sin una clara preponderancia de alguna IP en particular. En cambio la distribución del las IPs destino se presenta fragmentada en más areas grandes. Gran parte del espectro se divide de manera equitativa entre algunas pocas IPs. Por un lado la distribución de IPs origen presenta un 50 % dividido en pequeños pedazos, lo que indicaría un nivel de incertidumbre mayor a las "areas grandes" que presentan las IPs destino. Pero por otro lado estas últimas logran fragmentar, aunque sea en pocas IPs, lo que en el gráfico de IPs origen se presenta como un bloque de una única IP. Al parecer está pesando más este segundo análisis en el cálculo de la entropía, y por eso la de receptores es mayor.

Más adelante vemos que el hecho de que la cantidad de información de receptores es similar a la de emisores está intimamente relacionado con la función que cumplen los nodos en la red.

3.2.2. Análisis de capturas

Como se puede observar claramente en las figuras **??** y **??**, hay una gran presencia del dispositivo con la IP 10.10.99.213 cuyo comportamiento se limita a generar paquetes ARP y enviarlos, nunca es el destino de una consulta who-is. Por otro lado, a éste lo siguen dos dispositivos de igual cantidad de apariciones en los paquetes: las IPs 10.10.99.138 y 10.10.99.137. En las capturas, estos dos dispositivos se encuentran siempre como IPs destino, es decir, IPs por las cuales se consulta o se informa a qué MAC corresponde.

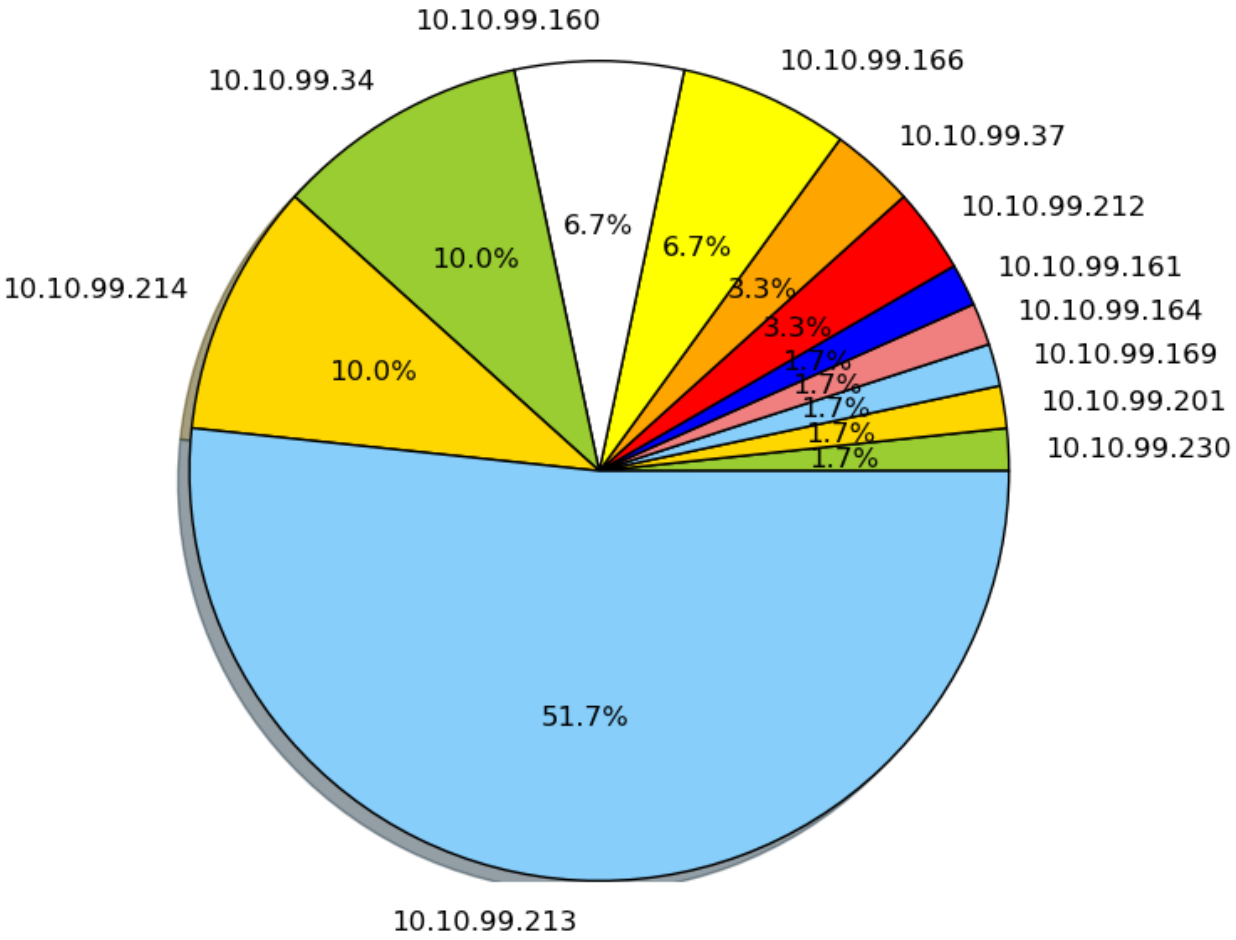

Figura 5: Proporción de IPs origen en los primeros 5 minutos de captura

En la figura 7 podemos ver la interacción que se corresponde con estos dispositivos. Vemos que casi la totalidad de los paquetes envíados por el dispositivo 213 están dirigidos a los dispositivos 137 y 138. Contrastando esto con los datos de los paquetes capturados vemos que el dispositivo 213 envía constantemente paquetes who-is preguntando por las IP 137 y 138, pero no se obtiene ninguna respuesta. Además de esto, el histograma nos muestra la interacción entre otros nodos de la red. El gráfico muestra al menos otros 5 nodos que son parte de esta interacción, lo interesante es que estos nodos no se conectan de manera centralizada.

Otros nodos que solo actuan de receptores de paquetes son el 35 y el 123. En la figura 7 y en el dump de capturas podemos ver que los que hacen peticiones sobre el nodo 135 son el nodo 34 y el 161. Y los que hacen peticiones sobre el 123 son el 212, 214 y 160.

A diferencia del experimento en los laboratorios del DC, vemos que en las oficinas de InvGate sí existen consultas entre nodos distintos, esto nos da una pauta de que el comportamiento de la red no está centralizado en un nodo particular, sino que está distribuido entre los nodos existentes.

Por otro lado es necesario aclarar que los nodos de los que nunca se obtiene respuesta, como el 25, 123, 137 y 138 probablemente sean dispositivos que no existen.

A raíz de estas observaciones, podemos concluir algunas cosas:

- O bien la red no provee de salida a Internet, o bien la salida a Internet está distribuida en diferentes nodos de la red, pues de no ser así veríamos una estructura más centralizada en un único nodo.
- Probablemente en algún momento hubo dispositivos en las IPs 10.10.99.25, 10.10.99.123, 10.10.99.137, 10.10.99.138, pero estos ya no forman parte de la red.
- Los dispositivos que aún pertenecen a la red, como el 213 o el 34 no parecen estar enterados que los

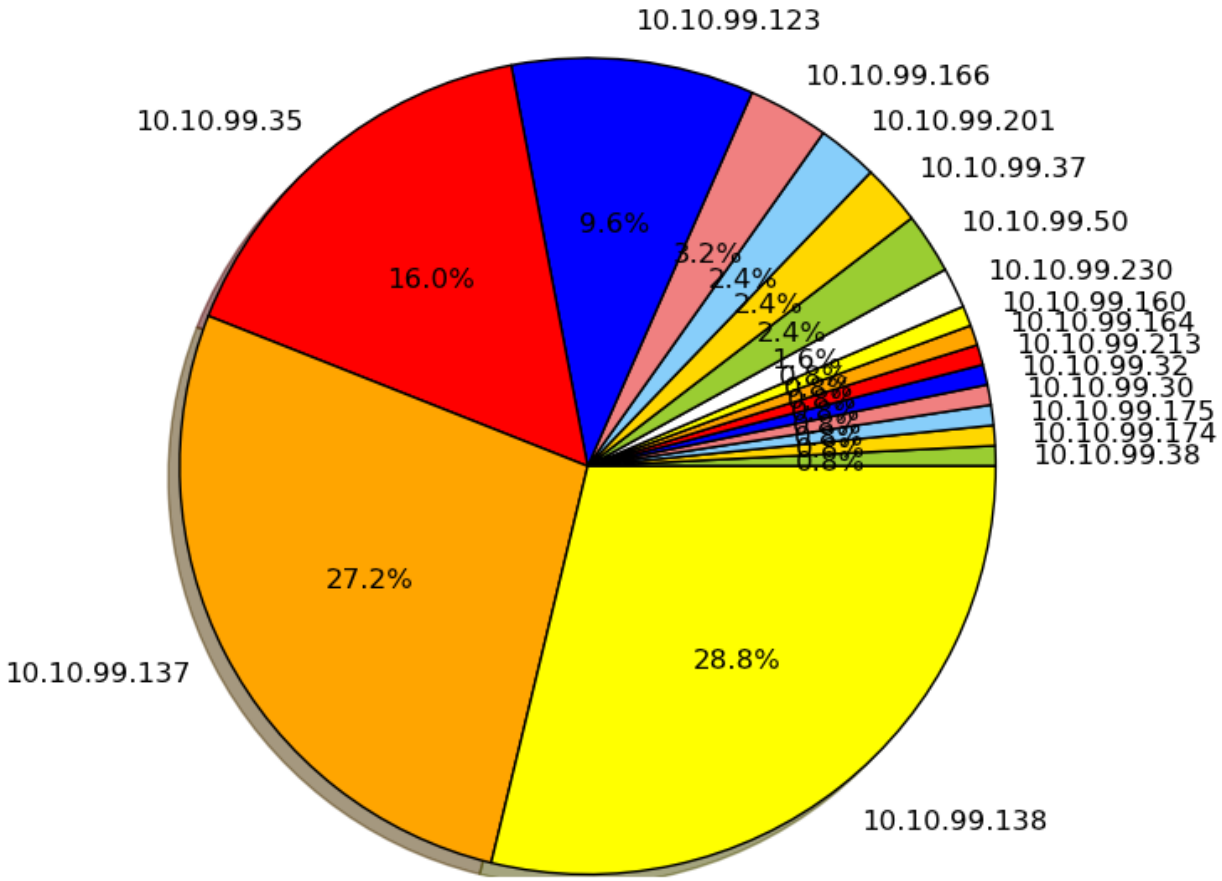

Figura 6: Proporción de IPs destino en los primeros 5 minutos de captura

otros dispositivos ya no existen.

- El dispositivo 213 parece estar configurado para conectarse con las IPs 10.10.99.137, 10.10.99.138, ya que constantemente intenta preguntar si algún dispositivo se encuentra en estas. Esto puede deberse por ejemplo a algun protocolo que quedó configurado en el dispositivo, que redirige el tráfico de cierto puerto a esas IPs.
- Todo esto nos lleva a pensar que la función de la red es más compleja que simplemente proveer Internet. Probablemente se trate de varios dispositivos que cumplen cada uno una función particular y se interrelacionan, como por ejemplo computadoras de trabajo, impresoras que reciben información de estas computadoras, dispositivos móviles.
- Esta hipótesis parece llevarse bien con el hecho de que la red elegida pertenece a un lugar de trabajo de desarrollo de software.

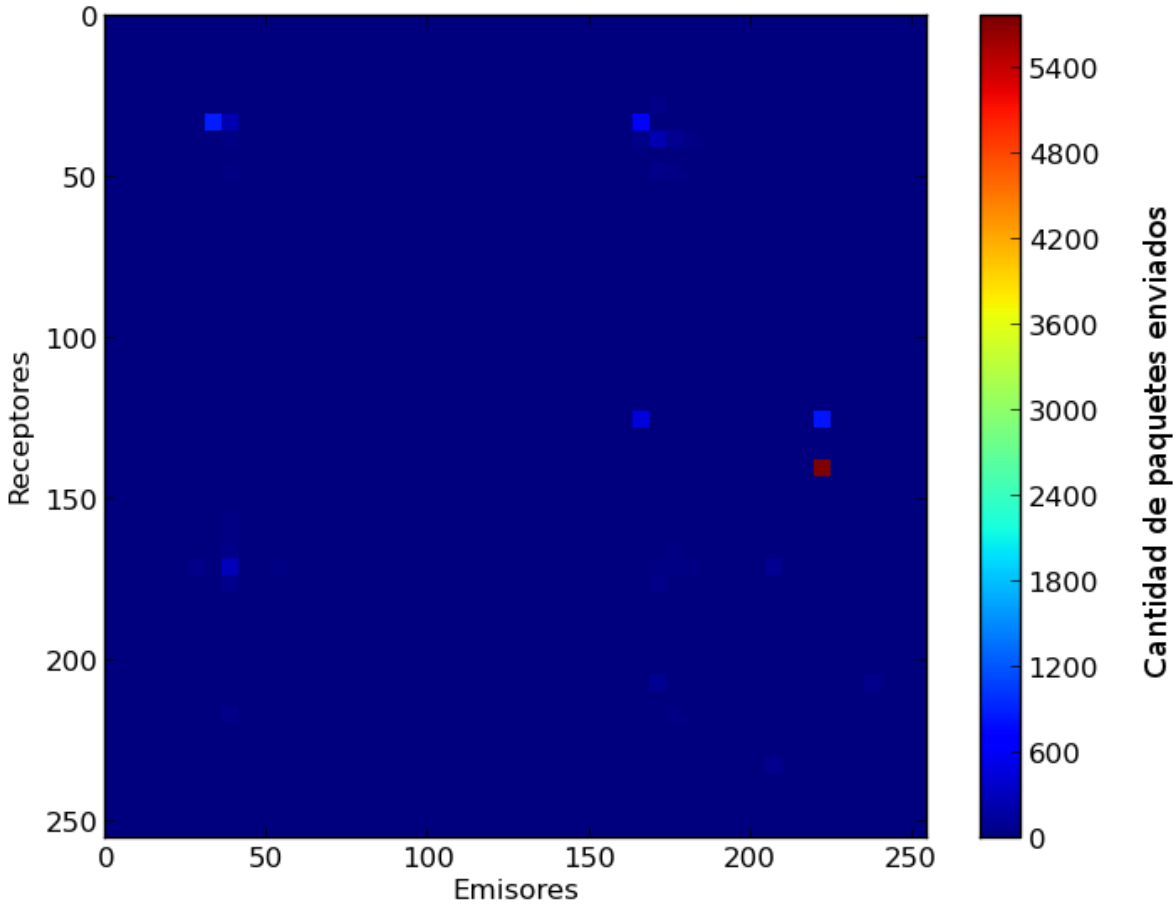

Figura 7: Relación entre emisores y receptores

4. Conclusiones

A lo largo del TP descubrimos que muchas cosas pueden averiguarse utilizando herramientas simples y al acceso de cualquiera con conocimientos básicos sobre redes o programación. Simplemente mirando los paquetes ARP (una mínima fracción de los totales) se puede descubrir los roles de los nodos, la cantidad promedio de dispositivos que se conectan por día, los horarios de mayor carga, la carga promedio del router y del servidor de DHCP, las funciones que cumple la red y cuánta comunicación hay entre dispositivos por fuera del router.

Todo esta información bien podría ser usada por un administrador de red para diagnosticar problemas como por un hacker con fines contrarios a los de la organización dueña de la red. En cualquier caso, el TP nos ayudó a tomar conciencia sobre la importancia de la seguridad en las redes, y en particular las públicas, dado que también aprendimos a usar herramientas que pueden obtener mucha más información de las que nos limitamos a coleccionar en este TP.

En particular, el uso de estas herramientas y la necesidad de sacar conclusiones a partir de las capturas nos hizo entender que no es dificil encontrarse en el día a día con distintos tipos de redes que tienen comportamientos y topologías diferentes. Y que estas diferencias se pueden inferir sin mucha dificultad tan solo a partir de muestras de las capturas del tráfico de los paquetes en la red.