Teoría de las Comunicaciones

Primer cuatrimestre 2014

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

Trabajo Práctico 1: Wiretapping

Grupo

Integrante	LU	Correo electrónico
Gastón Requeni	400/11	grequeni@hotmail.com
Sebastian Vita	149/11	sebastian_vita@yahoo.com.ar

Índice

1. Introducción

Este trabajo práctico consiste en escuchar pasivamente en una red de área local (LAN). Esto implica capturar paquetes que circulan en la red. Los paquetes pueden ser *unicast*, *broadcast* o *multicast*. Los primeros son los paquetes dirigidos a una MAC en particular (dirección de una interfaz de red de algún host en la red), los segundos son los dirigidos a todas las MACs en la red y los últimos son los dirigidos a un cierto grupo de MACs específico.

Nuestro objetivo será capturar los paquetes del protocolo ARP (Address Resolution Protocol). Este protocolo, según `RFC 826`, está diseñado para que todos los dispositivos en una red puedan encontrar la MAC address de una IP (asumiendo que se utiliza el protocolo IP a nivel de red). Cada vez que un dispositivo con dirección (`MAC1`,`IP1`) quiera enviar un paquete a la `IP2`, si no conoce la dirección física de esta IP, envía un paquete ARP broadcast, preguntando quién tiene la `IP2`. Luego se espera que el dispositivo con `IP2`, responda un mensaje unicast a `MAC1`, indicando `MAC2`.

A continuación vamos a establecer la notación para los paquetes ARP, utilizada en este informe:

MAC_SRC Dirección MAC Ethernet del host emisor.

MAC_DST Dirección MAC Ethernet del host receptor (dirección MAC a la cual se envía el paquete, pues el protocolo utiliza nivel de enlace para el envío).

ARP_OP Operación del paquete ARP. Puede ser `who-has` ó `is-at`.

ARP_MAC_SRC Dirección MAC del host emisor indicada en el paquete ARP.

ARP_IP_SRC Dirección IP del host emisor indicada en el paquete ARP.

ARP_MAC_DST Dirección MAC del host receptor indicada en el paquete ARP.

ARP_IP_DST Dirección IP del host receptor indicada en el paquete ARP.

Típicamente MAC_DST será desconocida y entonces se usará la dirección broadcast (`ff:ff:ff:ff:ff:ff`).

Cada host tendrá una tabla de traducciones de IP-MAC, de tal manera que cuando se desea enviar un paquete a una IP (a nivel de red), si corresponde a la LAN, se envía utilizando la MAC. Para llenar y actualizar esta tabla es que se utiliza ARP.

El algoritmo utilizado ante la captura de un paquete ARP indicado por `RFC 826` y asumiendo que siempre se utilizan protocolos Ethernet-IP, es el siguiente:

En el caso de que ARP_IP_SRC esté en nuestra tabla de traducciones, actualizamos la dirección de hardware asociada a ARP_IP_SRC. Para esto, reemplazamos el valor actual por ARP_MAC_SRC. Caso contrario, verificamos si ARP_IP_DST es nuestra IP. En cuyo caso, agregamos (ARP_IP_SRC,ARP_MAC_SCR) a la tabla de traducciones.

Por otro lado, si ARP_IP_DST es nuestra IP y ARP_OP=`who-has`:

1. *swap*(ARP_IP_DST, ARP_IP_SRC)
2. *swap*(ARP_MAC_DST, ARP_MAC_SRC)
3. *set*(ARP_MAC_SRC, nuestra dirección física)
4. *set*(ARP_OP, `is-at`)
5. Enviamos el paquete a MAC_DST = ARP_MAC_DST, respondiendo al host que nos haya enviado el paquete, y colocamos MAC_SRC = ARP_MAC_SRC.

Observar que la primera parte del algoritmo se realiza independientemente de la operación (ARP_OP). Esto permite que quien reciba un paquete `is-at` registre la dirección física y no envíe ninguna respuesta. Más aún, el algoritmo es muy laxo en cuanto a la validación de los campos del paquete, con lo cual podríamos establecer muchas variantes del funcionamiento típico mencionado previamente, que se acoplen a este algoritmo.

El objetivo del trabajo consiste en extraer propiedades características de una red utilizando la información provista por los paquetes ARP que circulan en la misma.

2. Desarrollo

Se realizará un análisis sobre 3 redes de área local distintas, intentando identificar características de las mismas en los paquetes del protocolo ARP.

Las redes elegidas fueron:

- Red Laboratorios DC: Esta red se sitúa en los laboratorios del departamento de computacion de la facultad de Ciencias Exactas y Naturales. Para poder acceder a la misma, tuvimos que loguearnos en una de las computadoras fijas del laboratorio. Se puede acceder también mediante una conexión wireless.
- Red Entrepiso: Esta red se encuentra en los pasillos de la Facultad de Ciencias Exactas y Naturales y en algunas oficinas. Se puede acceder por Ethernet o Wi-Fi, pero en este caso, accedimos por Wi-Fi para realizar las mediciones.
- Red Centro de Estudiantes (CECEN): Esta red se encuentra en el pabellón 2 de Ciudad Universitaria, en los alrededores del kiosko del centro de estudiantes de Exactas. Sólo se puede acceder de manera wireless.

El análisis comenzará tomando una muestra de paquetes capturados en cada una de las redes durante **30 minutos**. Luego usaremos una herramienta de software para leer esos datos y procesarlos de distintas maneras. En la secciones posteriores, desarrollaremos distintos métodos de análisis de los datos, y luego mostraremos y discutiremos los resultados.

2.1. IPs Más Solicitadas

De entre todos los paquetes de la muestra vamos a quedarnos únicamente con los que sean del tipo `who-has` y broadcast. El objetivo será buscar las IPs que más fueron solicitadas como destino del `who-has`, es decir las IPs por las que más se preguntó en la red.

Para esto vamos a contar la cantidad de veces que cada IP fue solicitada. Esta metodología tiene el inconveniente de que aparezca una IP con una cantidad de solicitudes inmensa en un intervalo de tiempo muy reducido. Eso corresponde a un sesgo en la medición, dado que consideramos que una IP es muy solicitada cuando es solicitada uniformemente a lo largo de los 30 minutos de la captura.

Para evitar en gran medida el sesgo, vamos a tomar intervalos de tiempo de tamaño n minutos (n divisor de 30). Luego vamos a tener en cuenta sólo a las IPs que hayan sido solicitadas en *todos* los intervalos al menos una vez, y contar las veces que se solicitaron éstas en los 30 minutos. La elección de n no debería ser arbitraria, es por eso que realizaremos algunos experimentos para determinar el valor de n que muestre los datos más apropiados, y mostraremos en este informe sólo algunos de ellos (los más significativos).

Este análisis se complementa con los grafos de relaciones ARP.

2.2. Grafos de relaciones ARP

2.3. Información y Entropía

Para el siguiente análisis consideraremos dos fuentes teóricas:

- $S_{src} = s_1, s_2, \ldots, s_n$: Cada símbolo s_i es una dirección IP que aparece en el campo ARP_IP_SRC de los paquetes ARP.
- $S_{dst} = d_1, d_2, \ldots, d_n$: Cada símbolo d_i es una dirección IP que aparece en el campo ARP_IP_DST de los paquetes ARP.

Por cada paquete ARP capturado, se considera que estas fuentes emitieron un símbolo (ARP_IP_SRC ó ARP_IP_DST respectivamente).

La idea es analizar la información de cada símbolo y compararla con la entropía de la fuente. ???

2.4. Paquetes destacados ("raros")

En las secciones posteriores estudiaremos algunos paquetes observados que resultaron extraños comparados con el uso estándar del protocolo ARP.

2.4.1. ARP_OP=who-has, broadcast, ARP_IP_SRC=0.0.0.0

Encontramos diversos paquetes del estilo:

MAC_SRC	MAC_DST	ARP_OP	ARP_MAC_SRC	ARP_IP_SRC	ARP_MAC_DST	ARP_IP_DST
00:27:0e:0d:f3:4b	ff:ff:ff:ff:ff:ff	who-has	00:27:0e:0d:f3:4b	0.0.0.0	00:00:00:00:00:00	10.2.5.14
¿Quién tiene la IP 10.2.5.14? Informar a 0.0.0.0 (00:27:0e:0d:f3:4b)						

El paquete fue capturado en "Red Labos DC". Es un paquete broadcast preguntando por una IP en particular, pero la IP fuente es nula. Buscamos otros paquetes correspondientes a la IP 10.2.5.14, y encontramos el siguiente:

MAC_SRC	MAC_DST	ARP_OP	ARP_MAC_SRC	ARP_IP_SRC	ARP_MAC_DST	ARP_IP_DST
00:27:0e:0d:f3:4b	ff:ff:ff:ff:ff:ff	who-has	00:27:0e:0d:f3:4b	10.2.5.14	00:00:00:00:00:00	10.2.5.249
¿Quién tiene la IP 10.2.5.249? Informar a 10.2.5.14 (00:27:0e:0d:f3:4b)						

Esto nos llamó mucho la atención, y luego de realizar una investigación al respecto, encontramos que estos paquetes se usan para detectar direcciones IP duplicadas[1][2]. Como se observa en el ejemplo, la IP 10.2.5.249 corresponde a la MAC address 00:27:0e:0d:f3:4b, que es la misma que envió el paquete extraño que mencionamos primero.

Si observamos el algoritmo de un host receptor de un paquete ARP (sección 1): Si el host receptor tiene la misma IP que el emisor, en el paso 2 responderá con un paquete que indique su MAC. Entonces el host emisor, al recibirlo, detectará que hay una IP duplicada.

2.4.2. ARP_OP=who-has, broadcast, ARP_IP_SRC=ARP_IP_DST (ARP gratuito)

Es una técnica utilizada para *anunciar* que un host es *dueño* de una IP. Los demás hosts de la red, cuando reciban este mensaje, pueden tenerlo en cuenta o no: Si alguien tiene la ip duplicada, puede modificarla para evitar que haya duplicados o simplemente ignorar el mensaje y seguir teniendo la ip duplicada, o tal vez enviar otro ARP gratuito (dar pelea...). En UNIX se respeta y acepta la ip de un host que envía un ARP gratuito. Ver [1].

Los paquetes tienen este formato: (ejemplo extraído de Red Labos DC)

MAC_SRC	MAC_DST	ARP_OP	ARP_MAC_SRC	ARP_IP_SRC	ARP_MAC_DST	ARP_IP_DST
b8:af:67:a1:ea:9e	ff:ff:ff:ff:ff:ff	who-has	b8:af:67:a1:ea:9e	10.2.0.187	00:00:00:00:00:00	10.2.0.187
b8:af:67:a1:ea:9e anuncia que tiene la IP 10.2.0.187						

Este formato también sirve para hacer ARP spoofing. Un host envía este paquete y, en una red de host UNIX, podría empezar a recibir los paquetes destinados a otra IP. A partir de eso, se podrían modificar los paquetes y reenviarlos al verdadero destinatario, o simplemente hacerce pasar por el destinatario y responder al emisor. Para más detalles sobre este uso, ver [3].

Existe otra manera de generar un ARP gratuito, que es enviando un paquete `is-at` broadcast, donde ARP_MAC_SRC = ARP_MAC_DST, ARP_IP_SRC = ARP_IP_DST (ver [1]). No observamos paquetes de este estilo en nuestras mediciones.

2.4.3. ARP_OP=who-has, unicast

2.4.4. ARP_OP=is-at, broadcast, no es ARP gratuito

2.4.5. ARP_OP=who-has, broadcast, ARP_IP_DST = ip pública

3. Resultados

3.1. Red Laboratorios DC

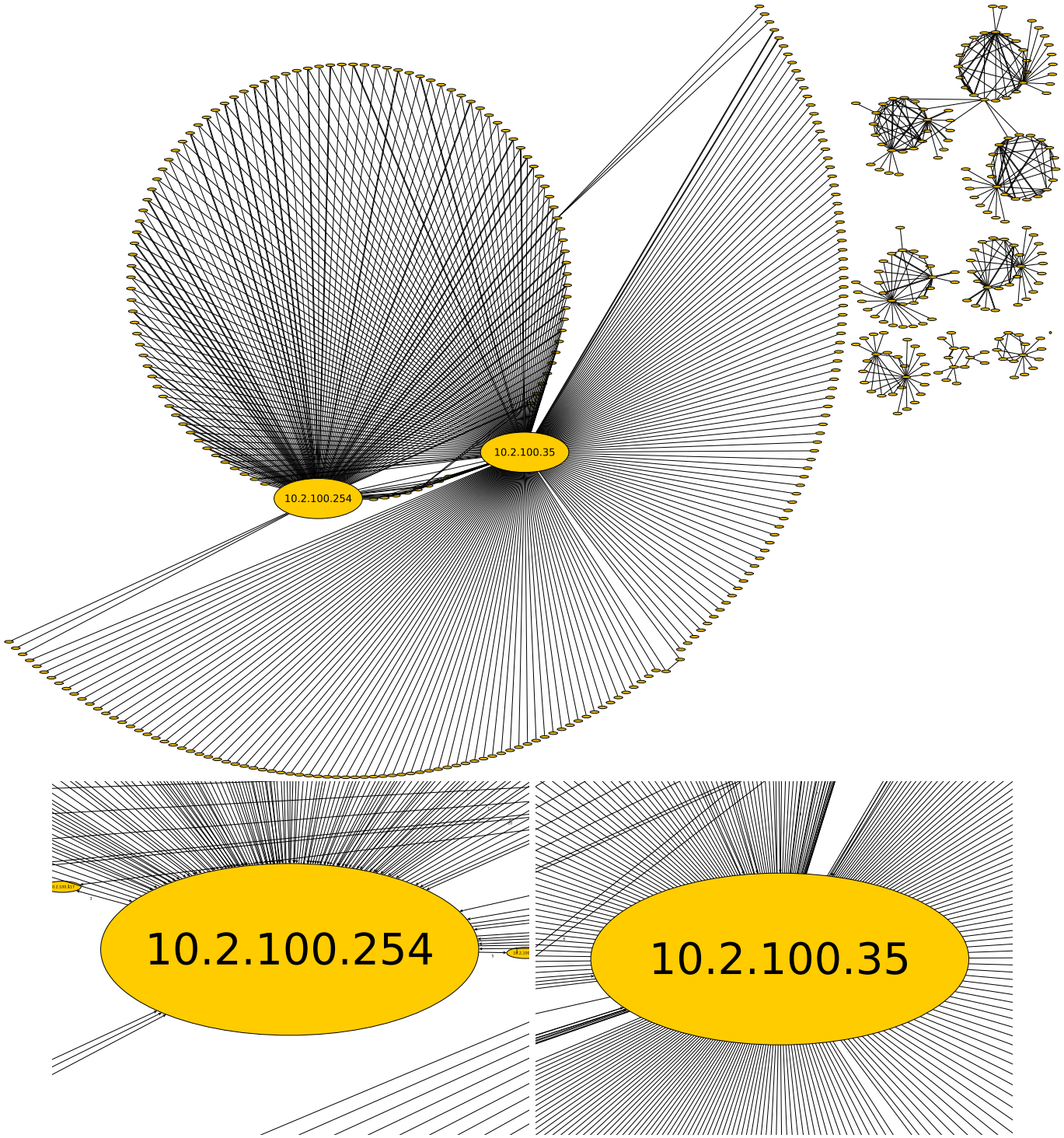

Figura 1: Envío de paquetes ARP en RedLabosDC durante media hora de muestreo. Los nodos corresponden a IPs. Los ejes indican un envío de paquete `who-has` broadcast, relacionando IP fuente con IP destino. El peso de los ejes es la cantidad de paquetes capturados. Ver zoom en Figura 2.

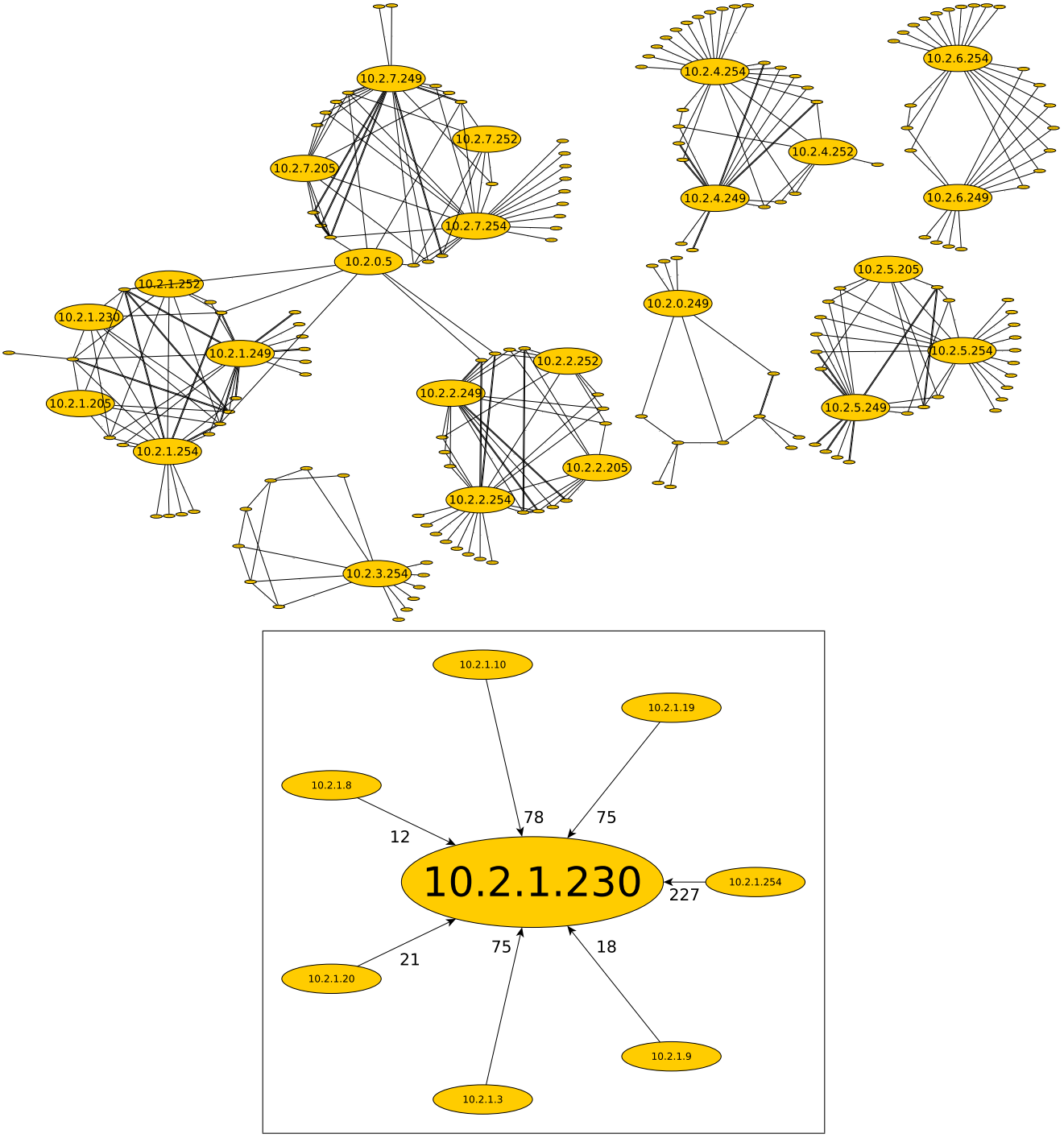

Figura 2: Envío de paquetes ARP en RedLabosDC durante media hora de muestreo. Los nodos corresponden a IPs. Los ejes indican un envío de paquete `who-has` broadcast, relacionando IP fuente con IP destino. El peso de los ejes es la cantidad de paquetes capturados. Abajo una ampliación de un sector del grafo superior.

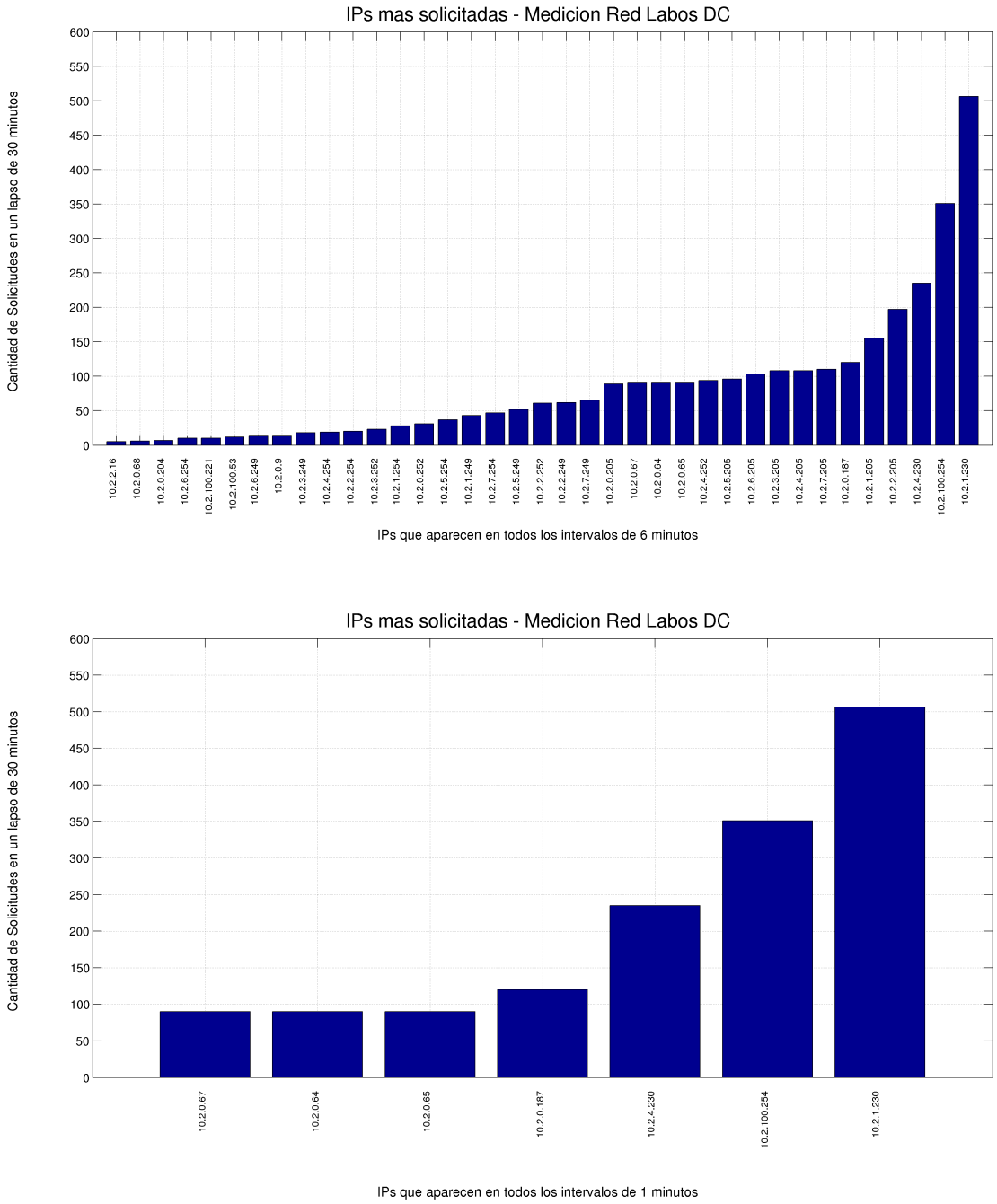

Figura 3: IPs más solicitadas - Red Labos DC - Intervalos de 6 minutos y de 1 minuto. Las IPs mostradas son las que recibieron `who-has` en todos los intervalos de tiempo de dicha longitud, de la medición de 30 minutos.

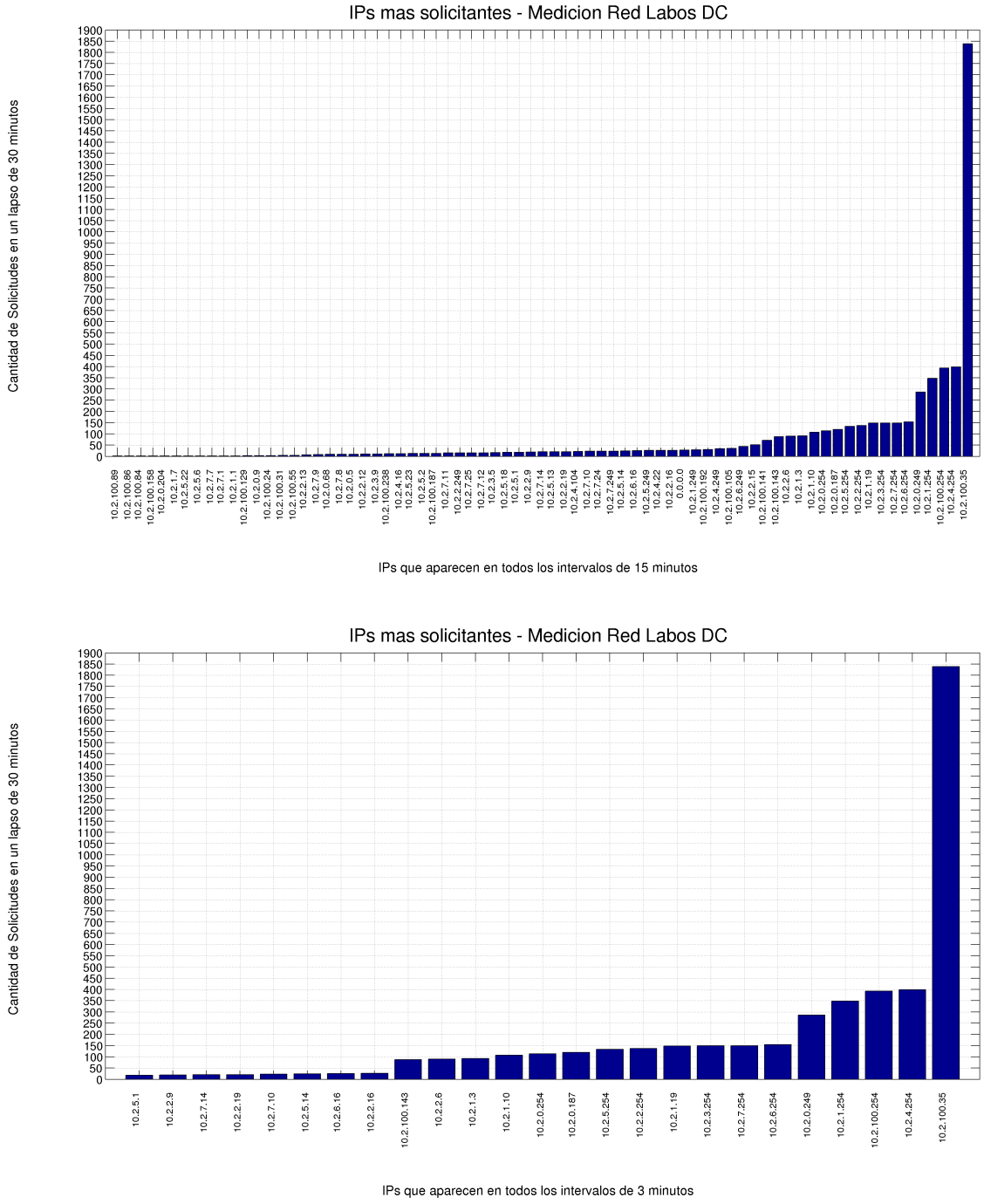

Figura 4: IPs más solicitantes - Red Labos DC - Intervalos de 15 minutos y de 3 minutos. Las IPs mostradas son las que enviaron `who-has` en todos los intervalos de tiempo de dicha longitud, de la medición de 30 minutos.

3.2. Red Entrepiso

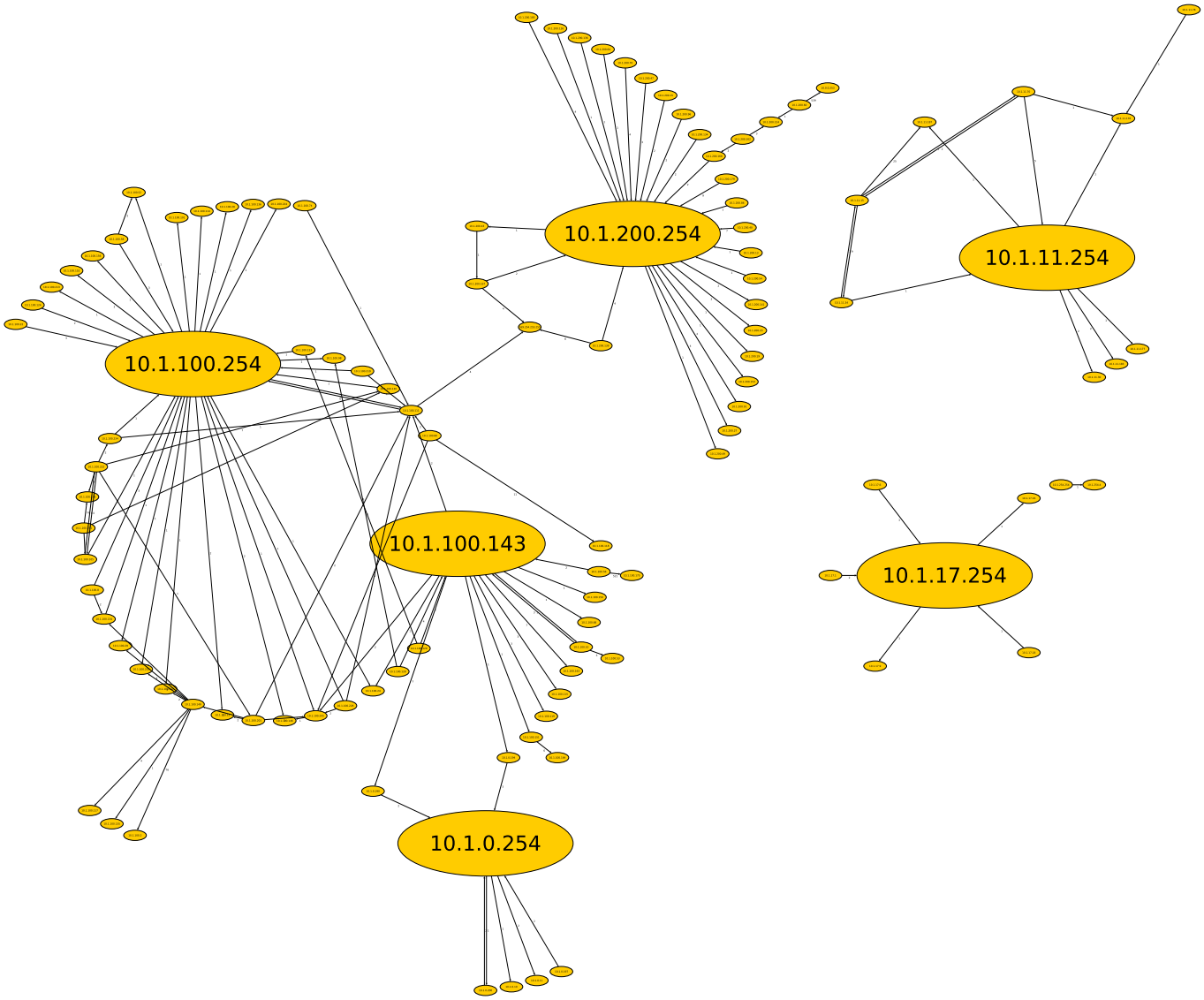

Figura 5: Envío de paquetes ARP en RedEntrepiso durante media hora de muestreo. Los nodos corresponden a IPs. Los ejes indican un envío de paquete `who-has` broadcast, relacionando IP fuente con IP destino. El peso de los ejes es la cantidad de paquetes capturados.

3.3. Red Centro de Estudiantes

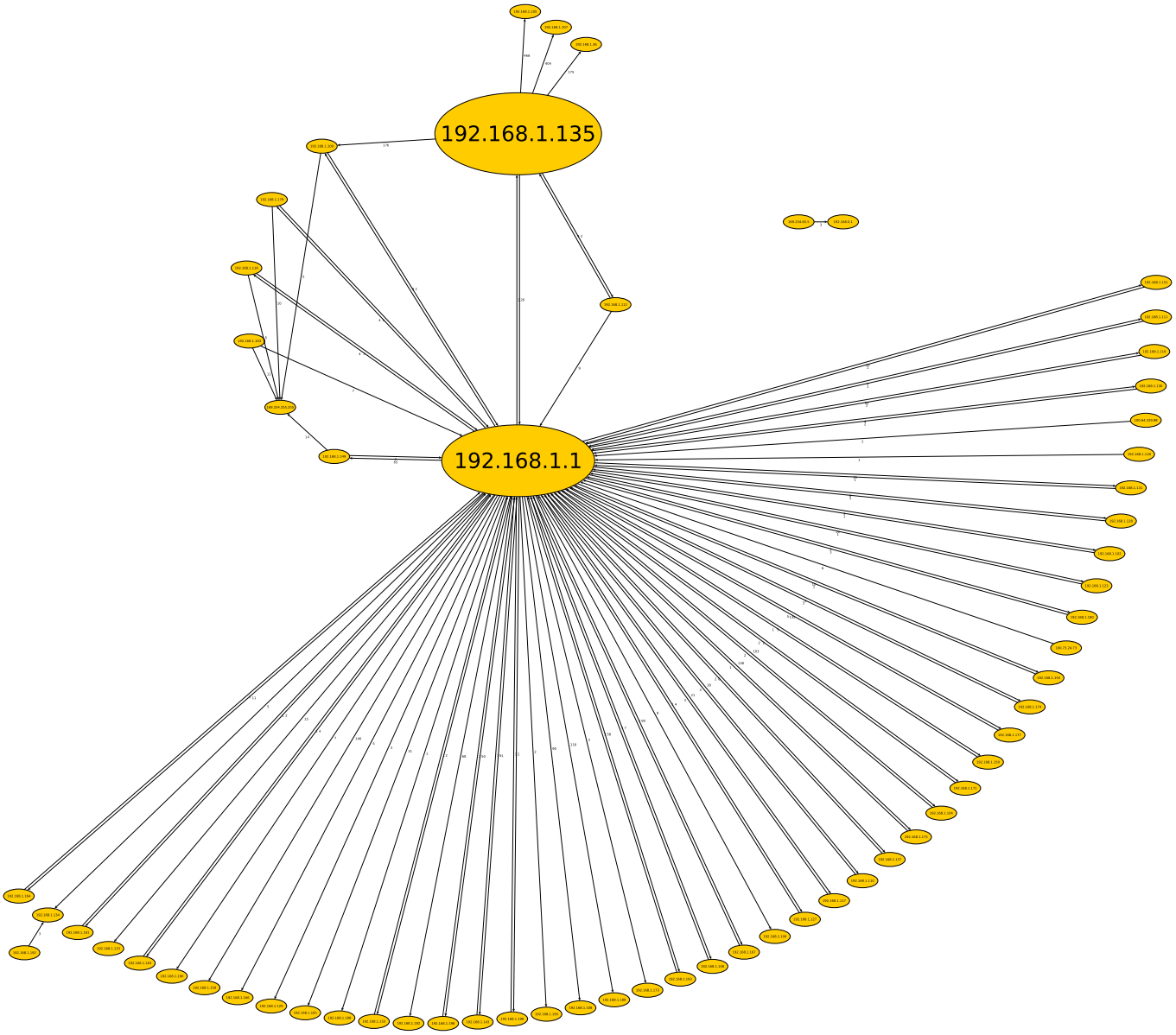

Figura 6: Envío de paquetes ARP en RedCECEN, durante media hora de muestreo. Los nodos corresponden a IPs. Los ejes indican un envío de paquete `who-has` broadcast, relacionando IP fuente con IP destino. El peso de los ejes es la cantidad de paquetes capturados.

4. Bibliografía

[1] http://linux-ip.net/html/ether-arp.html

[2] `man arping`

[3] http://ettercap.github.io/ettercap/, http://linux.die.net/man/8/ettercap (ver opción `arp` del programa).