DEPARTAMENTO
DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 1

ARP Sniffing

Teoría de las Comunicaciones
Primer Cuatrimestre de 2017

Integrante	LU	Correo electrónico
Borgna, Agustín	079/15	aborgna@dc.uba.ar
Gonzalez Benitez, Juan	324/14	gonzalezjuan.ab@gmail.com
Lancioni, Gian Franco	234/15	gianflancioni@gmail.com
Vazquez, Cristian	056/10	cristianvazquez4@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

Índice

A modo muy general, con el objetivo de analizar distintas redes locales, vamos a modelar ciertos comportamientos en el tráfico como fuentes de Teoría de la Información y poder extraer conclusiones concretas partiendo de las herramientas conocidas, particularmente los conceptos de entropía de la fuente e información de un símbolo en dicha fuente, para estudiar dichas fuentes.

1. Introducción

1.1. ARP

En particular nos interesa analizar las interacciones que se dan entre hosts de la red con el fin resolver direcciones de capa de red hacia capa de enlace, usualmente IP a MAC. Estas interacciones son propias del *ARP (protocolo de resolución de direcciones)* [1], el cual cumple un rol crucial en lo que conocemos como *internetworking*.

Se trata de un protocolo de *request* y *response*, implementados a partir de paquetes *'who-has'*, que preguntan a un dominio de broadcast por la *MAC address* de una IP particular, y paquetes *'is-at'* que responden de manera unicast al emisor del paquete 'who-has'.

El tipo del paquete se determina por el campo de operación de dicho paquete, del cual haremos amplio uso.

Por supuesto, dicho protocolo no se ejecuta cada vez que se quiera efectuar una transmisión, sino que los resultados se almacenan temporalmente una *cache* para agilizar tiempos.

1.2. Herramientas y resolución de primer etapa

Cada una de las capturas de red que hicimos (una por cada integrante del grupo), se hizo con el programa *TShark*, versión *terminal-based* del packet sniffer *Wireshark*.

Los paquetes obtenidos los manipulamos con el framework *Scapy* de *Python* que nos permite acceder a los campos de cada uno y, por ejemplo, determinar si tiene capa *ARP* y (de ser así) su destino buscado.

1.2.1. Primer ejercicio

Lo primero que se nos pide es modelar cada red como una fuete de información binaria S de memoria nula con símbolos en $\{S_{unicast},\ S_{broadcast}\}$.

Para implementar la fuente simplemente contaremos la cantidad de paquetes p tales que $p.dst =$ 'ff:ff:ff:ff:ff:ff' (i.e la dirección MAC de destino de broadcast). Dicha implementación está en *entro.py* y se encarga de imprimir por pantalla la entropía de la fuente y las ocurrencias de cada símbolo para una secuencia dada de paquetes.

Por lo tanto dicha fuente se abstrae de la identidad de los nodos de la red y los trata indiscriminadamente como emisores de símbolos según la dirección de capa de enlace de sus paquetes.

A modo analítico, la fuente S de cada red permite entender de qué manera se comunican los hosts de la red.

Por ejemplo, al tratarse de conexiones predominantemente dirigidas, resultaría extraño que los paquetes de broadcast fueran mayoría en el tráfico de la red. Y en caso contrario, la fuente haría visibles escenarios particulares de topologías o comunicaciones entre hosts.

1.2.2. Segundo ejercicio

El segundo ejercicio consistió en modelar una nueva fuente de información S_1 de memoria nula, con el objetivo de distinguir, en lugar de tipos de destinos como lo hacía S, los propios nodos (hosts) de la red.

Dicha distinción se hizo a partir de los paquetes *ARP*, y la implementación se encuentra en *distinguidos.py*.

Lo que hicimos fue considerar como símbolos las *IPs* de request de los mensajes 'who-has'. De esta manera, los símbolos con menos información son aquellos más solicitados.

Entonces los nodos distinguidos, con la finalidad de exponer aquellos más 'importantes' en la red, los consideramos como aquellos símbolos s tales que $I(s) < H(S_1)$ siendo $I(s)$ la información del nodo s en la fuente S_1 y $H(S_1)$ la entropía de la fuente.

Siguiendo con esta idea, es de esperarse que los nodos más solicitados, como los *default gateways* aparezcan siempre entre los distinguidos.

[1] https://en.wikipedia.org/wiki/Address_Resolution_Protocol

2. Mediciones realizadas

Ahora mostraremos los datos obtenidos en todas las mediciones, considerando las herramientas obtenidas en los dos ejercicios de la primera etapa mas un conjunto de gráficos que aporten nuevos datos para poder analizar.

2.1. Red 'Tatooine-lan'

Esta medición se trata de una red Wi-Fi doméstica de 4 dispositivos activos con un tiempo de escucha de 20 minutos. Dos de las máquinas estaban usando servicios de streaming de video.

Mientras se realizaba la medición también se borraron los registros (comando 'arp -d $< host\_ip >$' en linux) la caché ARP en el dispositivo cuya IP pública es 190.168.0.14 y privada 192.168.0.12 con el objetivo de forzar un handshake en dicho protocolo con el default gateway de la red.

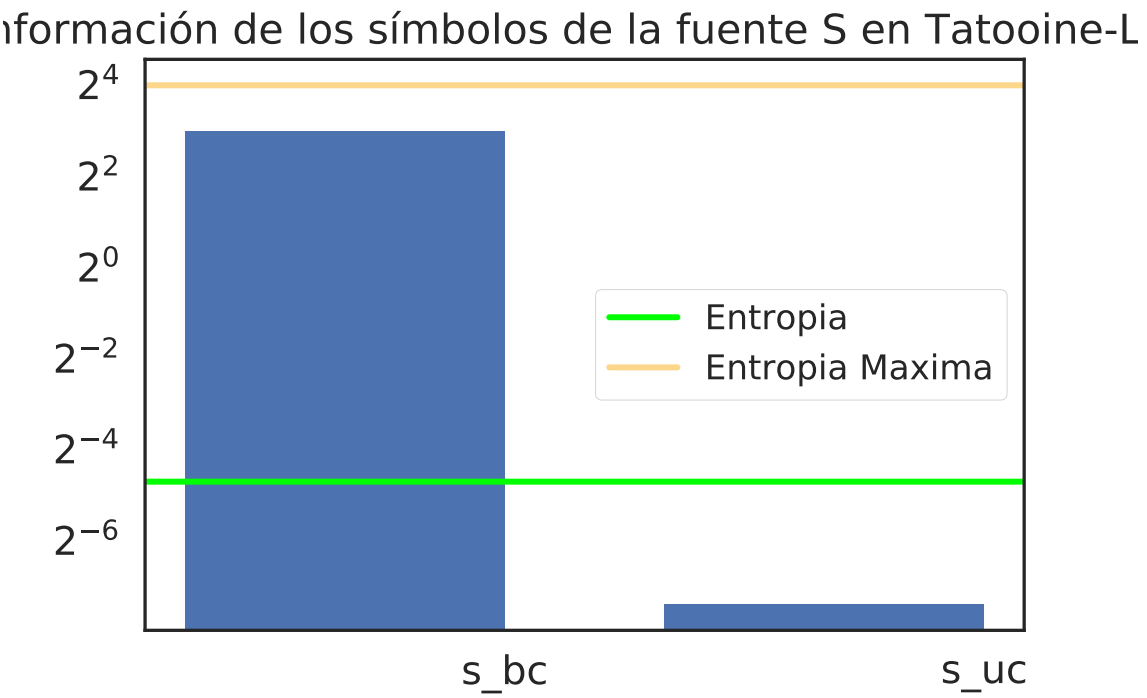
Además, se reinició un dispositivo mientras se corría la escucha.

2.1.1. Fuente S

En la medición se contaron entre todos los paquetes del archivo *.pcap* con la herramienta el ejercicio 1:

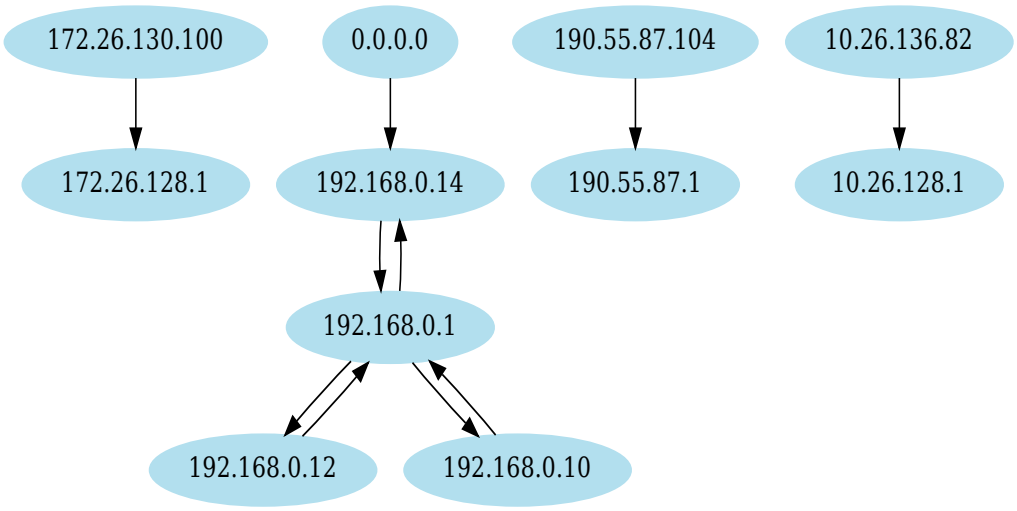
- 69834 paquetes unicast
- 289 paquetes broadcast (p.dst = 'ff:ff:ff:ff:ff')
- 70123 paquetes en total

Dejando una entropía $H(S) = 0{,}03858550573957047$ y la siguiente distribución en la información de los símbolos $S_{broadcast}$ y $S_{unicast}$:



2.1.2. Grafo de conectividad de la red

Con el fin de analizar de mejor manera la topología de la red también armamos un grafo entre los nodos de la red que muestre el tráfico de paquetes ARP sensado. Cada nodo está asociado a una IP y un eje entre una IP_1 y IP_2 significa que se sensó un paquete *'who-has'* emitido por IP_1 consultando por la dirección MAC de IP_2.



Se puede notar que ninguno de los otros dispositivos mencionados, al margen de aquel para el cual flusheamos la caché ARP, hizo un request de tipo *'who-has'* durante el sniff usando su IP pública. Dicho comportamiento se repite en las otras redes aledañas en las cuales un único nodo envía un paquete *'who-has'* a una IP con el último octeto en 1 (candidatos a routers).

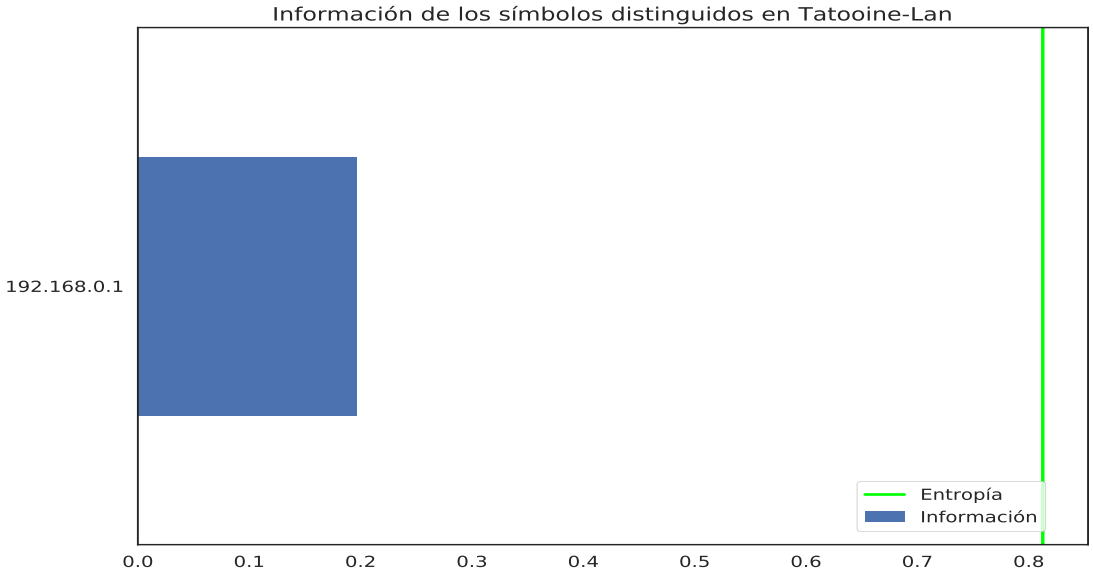
Pero, por otro lado, se puede ver buena actividad en una red que parece corresponderse con el rango 192.168.0.0/16 reservado para IPs privadas. Como la capa de enlace se encarga de que las redes privadas no se solapen, podemos afirmar que se trata de la red a la que nos conectamos durante la medición.

Es interesante la aparición de una IP 0.0.0.0, la cual se asigna a dispositivos en diferentes situaciones. Entre ellos, cuando un dispositivo se inicializa en una red y envía su primer paquete DHCP (sin conocer todavía su IP en la red). Por lo tanto, tiene sentido que dicha interacción se deba al dispositivo reiniciado mencionado antes.

2.1.3. Información y nodos distinguidos

Haciendo uso de la herramienta del ejercicio 2 para distinguir nodos de la red respecto de sus respectivos símbolos en la fuente S_1 en caso de que aporten menor información que la entropía de dicha fuente, encontramos únicamente al nodo 192.168.0.1 como nodo distinguido.

Las direcciones con los dos últimos octetos en 0.1 en redes 192.168.0.0/16 son comunes para IPs de interfaces de routers. Lo que significaría que nuestro código fue capaz de encontrar al menos un default gateway, particularmente por el que más nodos consultaron.



2.2. Red de una empresa

Este experimento sobre una red Wi-Fi en una empresa con alrededor de 15 empleados. En esta red se suelen conectar varios dispositivos móviles y también notebooks.

2.2.1. Fuente S

Utilizando las herramientas presentadas en el ejercicio 1 sobre esta red se obtuvieron los siguientes resultados

- 12039 paquetes unicasts
- 6561 paquetes broadcasts
- 18600 paquetes en total

La *entropía* para esta fuente se calculó en $0{,}936493026389$. El siguiente gráfico muestra este comportamiento.

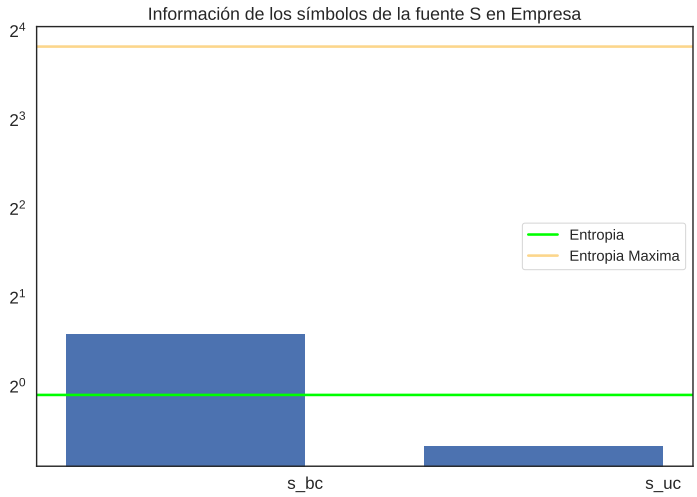


Figura 1: Entropía de la red

Como la medición se realiza sobre una red inalámbrica a la cual se encuentran conectados varios dispositivos, es razonable que exista una mayor presencia de paquetes *unicast*. Esto es porque dichos dispositivos, generalmente, envían paquetes a un destino IP particular con la finalidad de conectarse a algún servidor externo.

2.2.2. Grafo de conectividad de la red

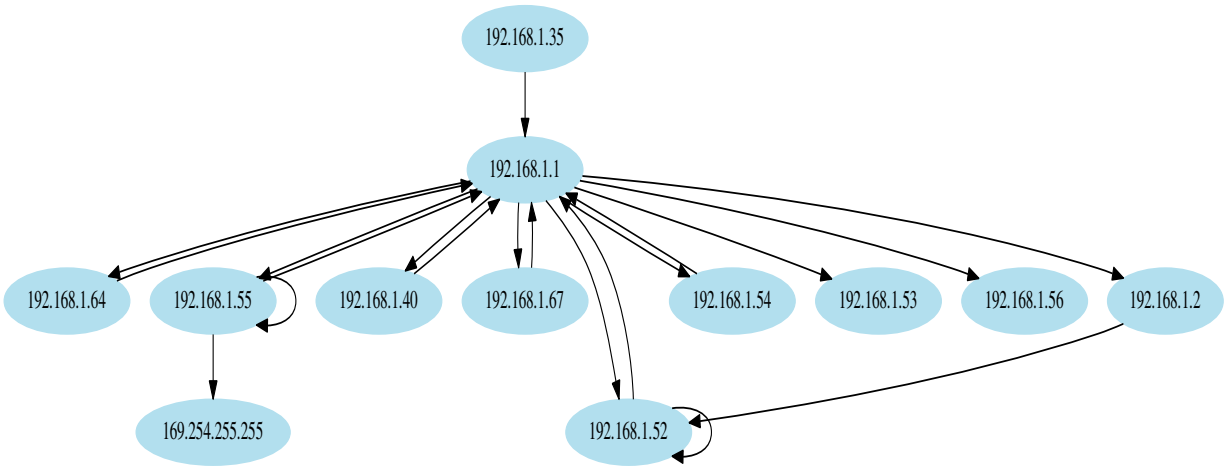


Figura 2: Grafo de conectividad

En la figura anterior se puede observar que el nodo 192.168.1.1 se comporta como un nodo raíz al cual se conectan varios dispositivos, este comportamiento es esperado debido a que se trata de una red Wi-Fi en el cual el nodo 192.168.1.1 podría ser el Router.

A su vez, se puede observar que las direcciones asignadas son de la forma 192.168.1.0 con una máscara $/25$ ya que ninguno de los *hosts* tiene asignado direcciones superiores a 126.

También se puede observar un pedido de asignación de IP por parte del host 192.168.1.55 hacia el servidor DHCP (169.254.255.255).

Por último podemos observar paquetes en las cuales la dirección de destino coincide con la dirección del emisor, esto se debe a que son *gratuitous ARP packets* los cuales suponemos son utilizados para anunciar a la red de la presencia de un nuevo dispositivo o para actualizar las tablas de ARP luego de que una MAC address cambia de IP.

2.2.3. Información y nodos distinguidos

Para esta sección reutilizamos la herramienta presentada en el ejercicio 2 con el objetivo de analizar la presencia de nodos distinguidos de la fuente.

Como se puede observar en la figura 3, el dispositivo 192.168.1.1 se comporta como único nodo distinguido de nuestra red y su dirección se corresponde con las asignadas generalmente por los routers, por este motivo no es precipitado suponer que podría tratarse del *default gateway* de esta red.

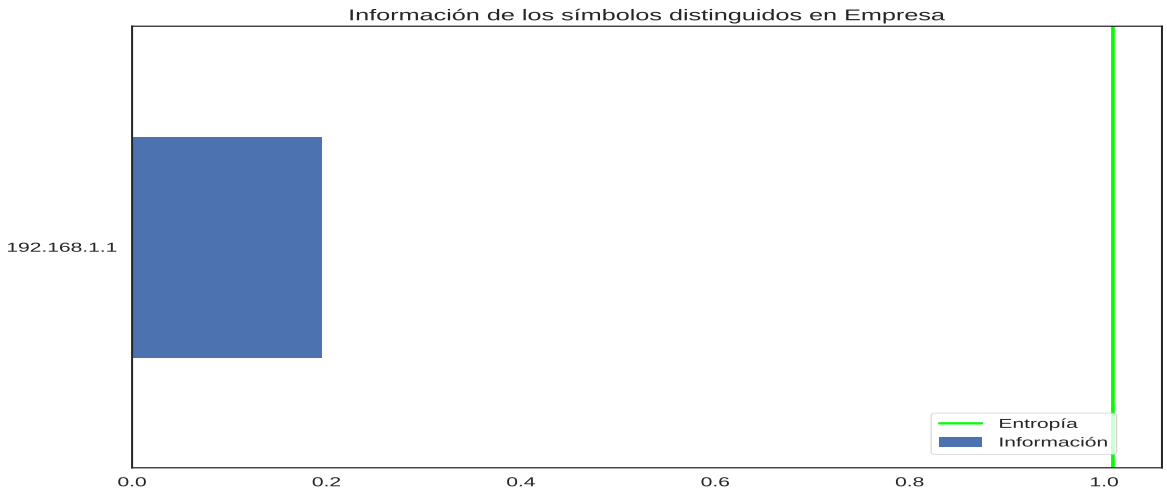


Figura 3: Nodos distinguidos

2.3. Red de una universidad

Este experimento se realizó sobre la red Wi-Fi de acceso público *EXACTAS-UBA*, en la zona de aulas del pabellón 1 de la FCEyN, realizando una captura de 45 minutos a las 19 horas. Suponemos que esta red es usada por alumnos y profesores durante las clases.

Durante la medición vació manualmente la caché ARP de un host.

2.3.1. Fuente S

Mediante las herramientas ya presentadas obtuvimos los siguientes resultados

- 38507 paquetes unicasts
- 103 paquetes broadcasts
- 38610 paquetes en total
- $H(S) = 0{,}02665$

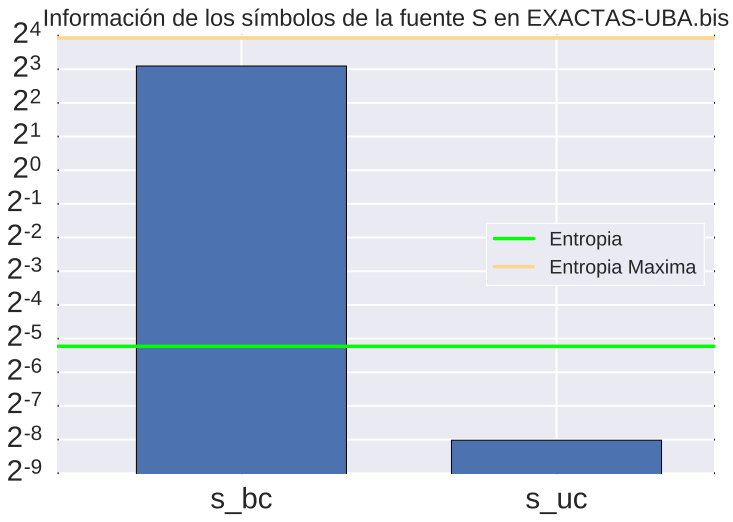


Figura 4: Entropía de la red

Las mediciones de la figura 4 se corresponden con el uso de la red puramente para acceder a servidores externos que se esperaría de los alumnos durante una clase. Efectivamente, analizando el tráfico comprobamos que el 99.7 % de los paquetes tienen como MAC de fuente o destino al default gateway.

2.3.2. Grafo de conectividad de la red

Grafo de conectividad de la red para visualizar el comportamiento de la misma.

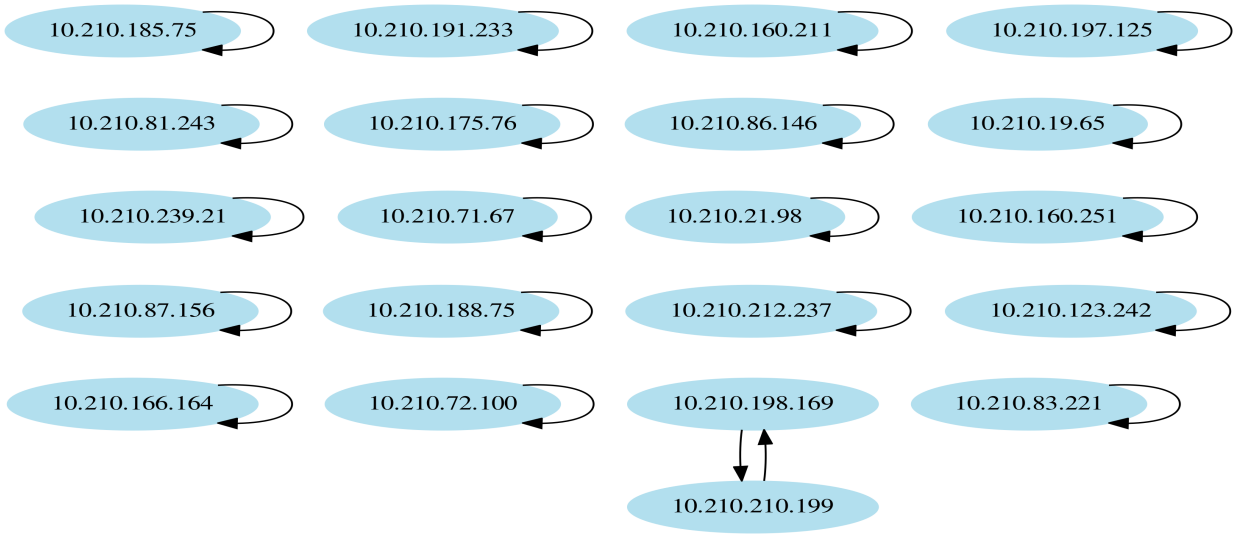


Figura 5: Grafo de conectividad

Al graficar el tráfico ARP sensado vemos que aparte de la petición por la MAC del gateway que se generó al vaciar la caché el resto del tráfico se compone enteramente de gratuitous ARPs. Esto se puede deber a que todos los hosts de la red ya estaban conectados cuando se inició la captura, y ya tenían en su caché la MAC del gateway.

2.3.3. Información y nodos distinguidos

Para esta sección reutilizamos la herramienta presentada en el ejercicio 2 con el objetivo de analizar la presencia de nodos distinguidos de la fuente.

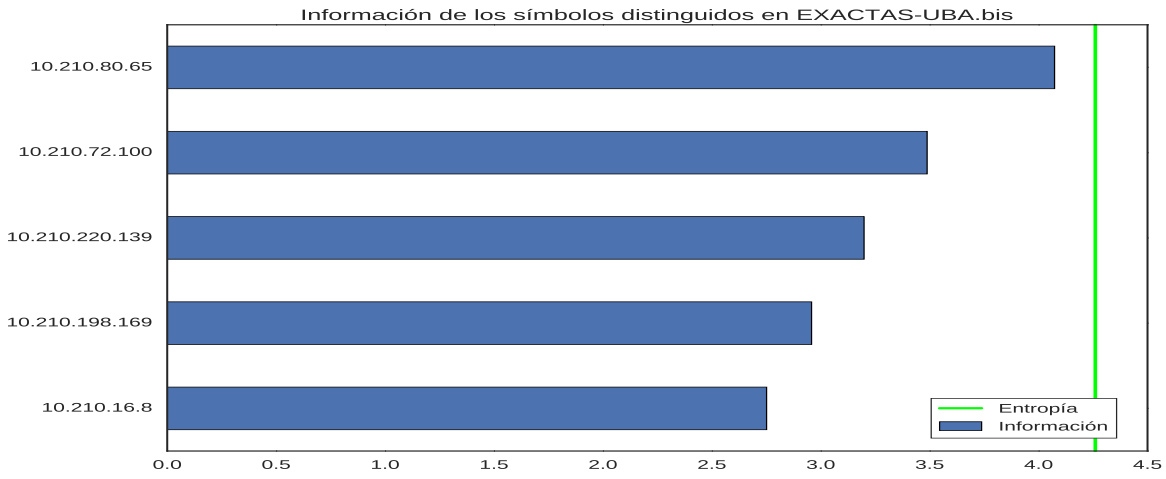


Figura 6: Nodos distinguidos

Como se puede observar en la figura 6, la gran cantidad de gratuitous ARPs no nos permite encontrar conclusivamente el default gateway.

2.4. Red de una universidad cableada

Este experimento se realizó sobre una red cableada del Departamento de computación, realizando una captura de 60 minutos.

Durante la medición vació manualmente la caché ARP de un host.

2.4.1. Fuente S

Mediante las herramientas ya presentadas obtuvimos los siguientes resultados

- 398 paquetes unicasts
- 1768 paquetes broadcasts
- 2166 paquetes en total
- $H(S) = 0{,}688$

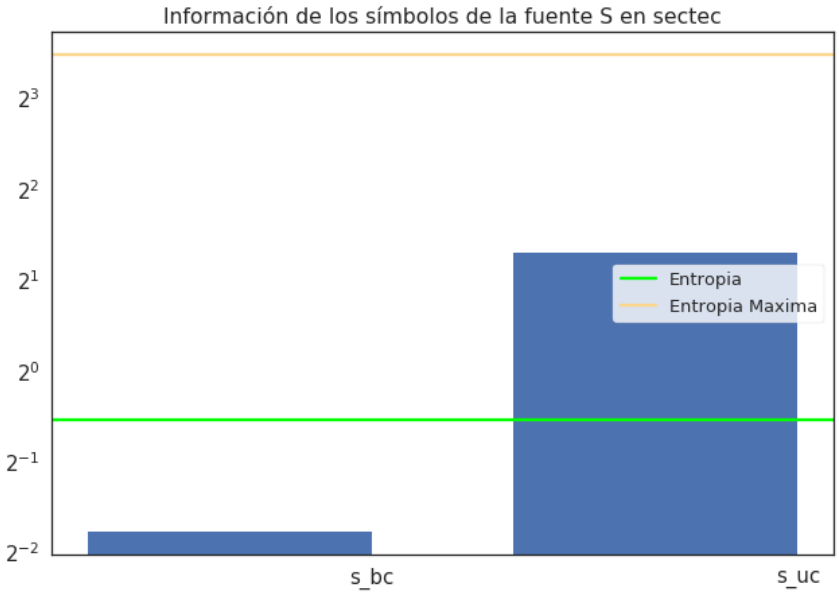


Figura 7: Entropía de la red

2.4.2. Grafo de conectividad de la red

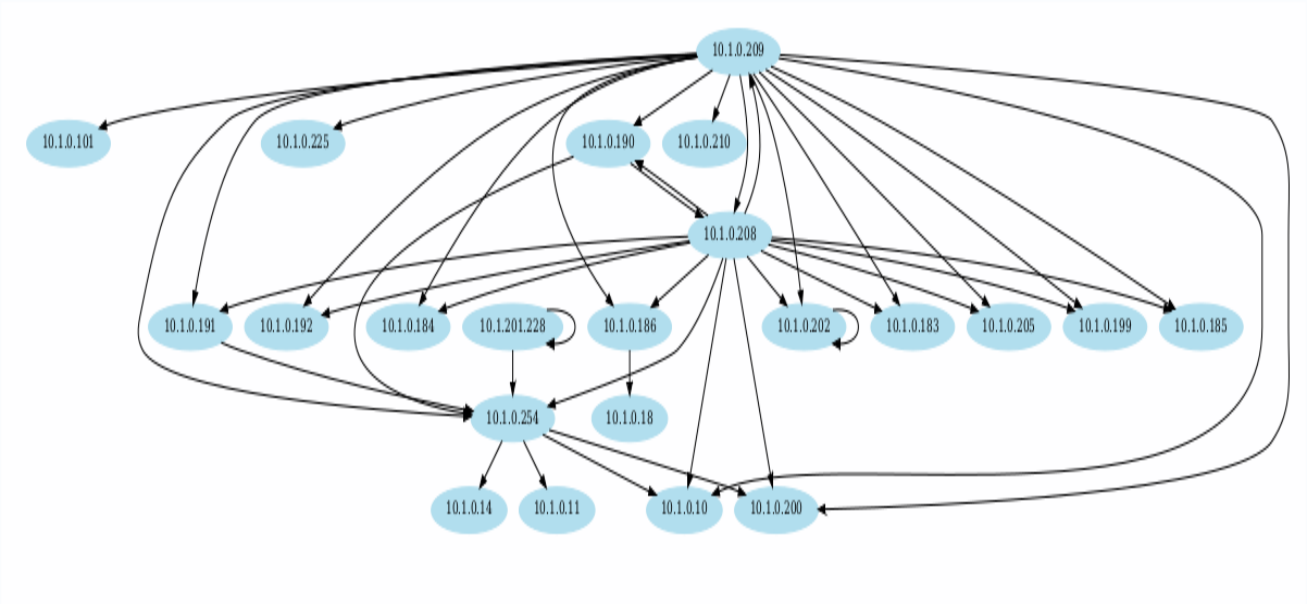


Figura 8: Grafo de conectividad

Podemos observar en la figura 8 como desde las ips *10.1.0.209* y *10.1.0.208* se emiten muchos broadcast ya que son *Access Point*.

El nodo que más conexiones recibe es *10.1.0.254* y por lo tanto deducimos que es el *Default Gateway* de la red.

2.4.3. Información y nodos distinguidos

Para esta sección reutilizamos la herramienta presentada en el ejercicio 2 con el objetivo de analizar la presencia de nodos distinguidos de la fuente.

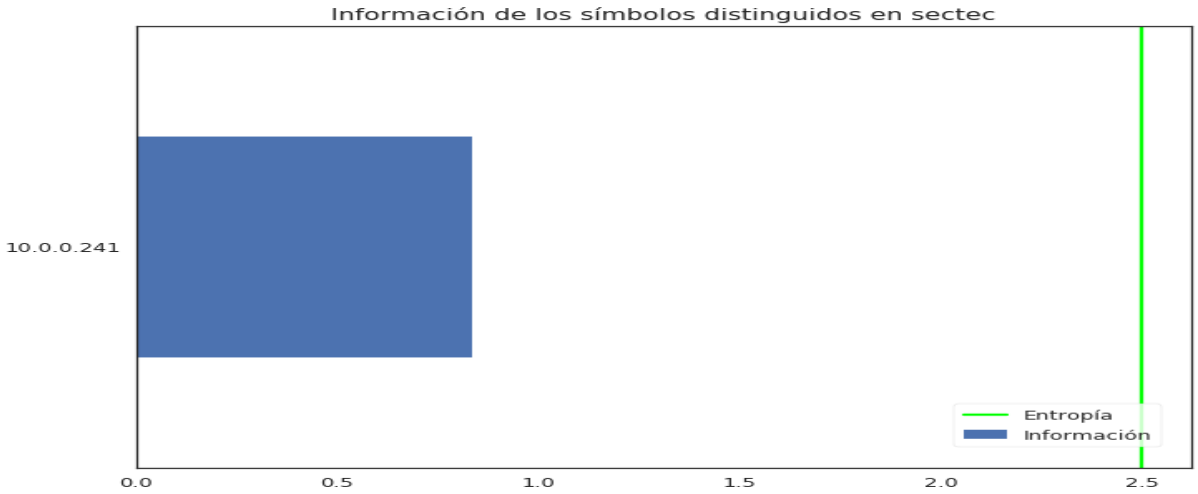


Figura 9: Nodos distinguidos

Notamos que el nodo distinguido no está en la misma red que nuestra medición. Consultamos y averiguamos que en esa dirección se entra el servidor de dhcp.