DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico I

Wiretapping

Teoría de las Comunicaciones
Segundo Cuatrimestre de 2015

Integrante	LU	Correo electrónico
Alvarez, Matías	090/12	matyy.alvarez@gmail.com
Dorr, Francisco	434/09	fran.dorr@gmail.com
Litwak, Brian	241/12	brian.litwak@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

1. Objetivo del trabajo

El objetivo de este trabajo es utilizar técnicas provistas por la teoría de la información para distinguir diversos aspectos de la red de manera analítica. Para ello, es sugerido el uso de dos herramientas modernas de manipulación y análisis de paquetes: Wireshark y Scapy.

1.1. Breve introducción

En primer lugar implementamos herramientas para simular fuentes de información. Una fuente de información, desde el punto de vista de la teoría de la información de Shannon, es un objeto que emite símbolos s_i, en donde cada uno tiene una probabilidad p_i de ser emitido. Entre ellas tendremos:

- Una herramienta que simula una fuente que emite paquetes Ethernet, es decir cada símbolo en este caso será un posible frame generado con dicho protocolo.
- Otra herramienta que sumila una fuente que emite paquetes que utilizan el protocolo ARP.

2. Presentación e hipótesis sobre los experimentos

Dividiremos la experimentación del trabajo en tres secciones, en donde cada una consistirá en dejar escuchando las herramientas en tres redes distintas. Entre ellas estarán las siguientes:

1. Red hogareña: consistirá en una pequeña red en donde tenemos de antemano la noción de cuántos equipos pueden estar conectados. En este caso, al ser tres personas las que conviven, sabemos que como máximo se pueden encontrar 7 equipos distintos, entre ellos 3 celulares, 2 tablets, una notebook y un televisor. La duración de la prueba será de aproximadamente 6 horas para tener mayor cantidad de información, ya que sabemos que la cantidad de equipos que se puede conectar es baja.
2. Red del laboratorio de computación: en este caso llevaremos una notebook a la universidad para observar qué cantidad de equipos se conectan en un determinado período de tiempo. Sabemos de antemano que entre las 17 y 22hs los laboratorios suelen estar llenos por alumnos de la carrera de computación, por lo cual creímos que sería interesante observar lo que sucede en horarios en donde dichos alumnos no acaparan los laboratorios por estar en clase, sino por juntarse para resolver trabajos prácticos, resolver prácticas, etc. Es por esto que la medición se realizó entre las 13 y 14hs.
3. –

Antes de realizar la experimentación plantearemos algunas hipótesis sobre los resultados.

- En general creemos que la dirección MAC de broadcast (FF:FF:FF:FF:FF:FF) será la más pedida, ya que es la dirección en donde los hosts pueden mandar mensajes a todos los demás equipos.
- De la misma forma, en general, creemos que habrá una cantidad similar de protocolos en todas las redes ya que son redes públicas a las que acceden la mayoría de los equipos cotidianos como ser computadoras, celulares y tablets.
- En la red hogareña creemos que habrá poca cantidad de envío de paquetes ARP *who is* dado que la cantidad de dispositivos es muy acotada.
- En la red del laboratorio creemos que esto último será al revés, es decir, habrá mucho envío de paquetes ARP *who is* de distintas MAC, pues la mayor cantidad de conexiones suele provenir de celulares y éstos tienden a bloquearse y desbloquearse cada poco tiempo (y cada vez que se realiza esa acción, la gran mayoría de los celulares apaga el wifi de a momentos para ahorrar batería). Por lo cual cada reconexión implica una nueva tanda de mensajes ARP y es por esto que creemos que la cantidad final será muy elevada.
- –

Las hipótesis generales serán corroboradas en el final del informe y las que dependen de cada red, al final de cada experimentación.

3. Experimentación

3.1. Red hogareña

Comenzaremos la experimentación recolectando datos sobre la cantidad y diversidad de protocolos.

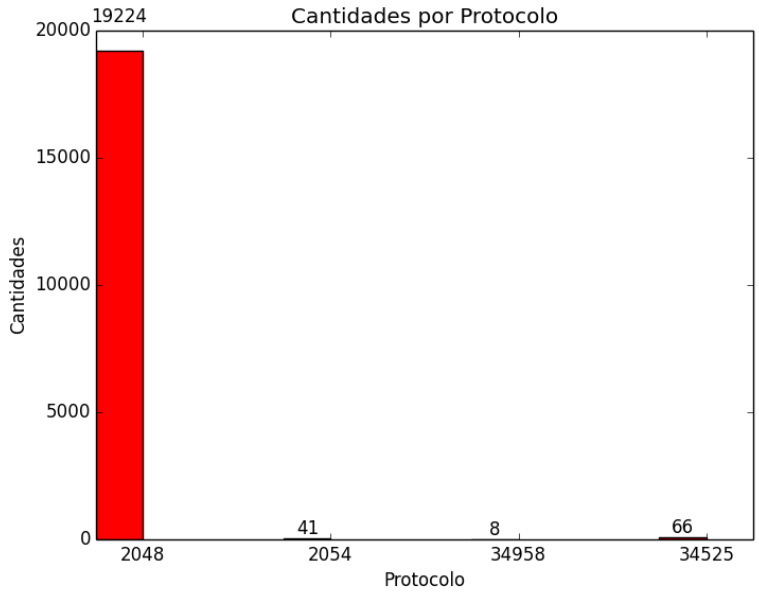


Figura 1: Distintos tipos de protocolos (EtherType) encontrados en la red, organizados por cantidad

En primer lugar vemos que se cuenta con una pequeña cantidad de protocolos y cuando buscamos en la lista de EtherTypes de IEEE[1] encontramos que el código 2048 pertenece a mensajes de tipo *IPv4*, el 2054 a mensajes *ARP*, el 34958 a lo que se denomina como *Portbased network access protocol* y el 34983 a *Service VLAN tag identifier*.

El gráfico resulta bastante intuitivo ya que la mayoría de los mensajes que emite un dispositivo cotidiano como ser un celular o una computadora suelen ser paquetes IP que viajan por internet para realizar pedidos, por lo cual son de tipo IPv4 (que es el modo utilizado en la actualidad) y esto explica la gran cantidad de estos paquetes. Por otro lado, la proporción de ARP resulta muy baja debido a la cantidad de dispositivos que existen en la red. Los otros protocolos, creemos que son necesarios por el router de la casa por la configuración del proveedor de servicios.

[1] http://www.iana.org/assignments/ieee-802-numbers/ieee-802-numbers.xhtml

Veamos a continuación el mismo gráfico detallando las probabilidades de cada símbolo (protocolo).

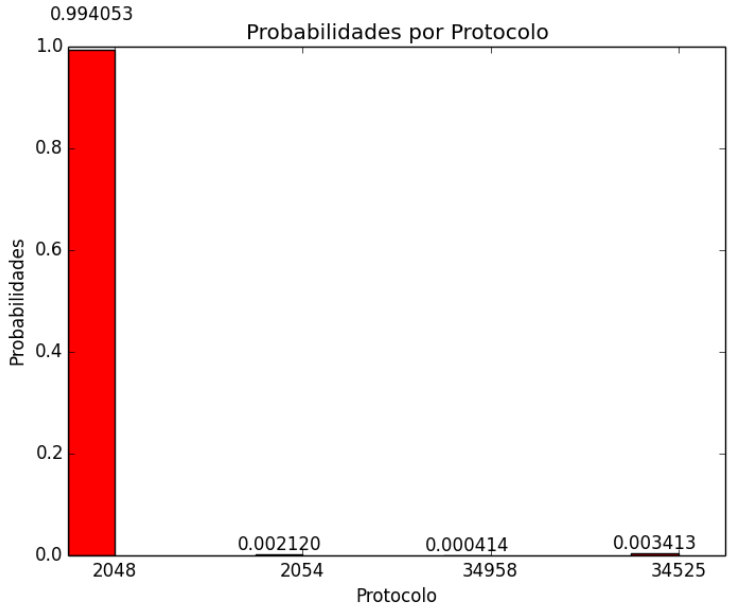


Figura 2: Probabilidad por cada tipo de protocolo detectado

La proporción de paquetes ARP por sobre el resto de los paquetes termina siendo del 0.0021 % , lo que en principio podría indicarnos que tiene una fuerte relación con la baja cantidad de dispositivos (es decir, lo que ya sabemos desde la teoría).

Veamos a continuación el gráfico de la entropía a medida que se transmiten paquetes.

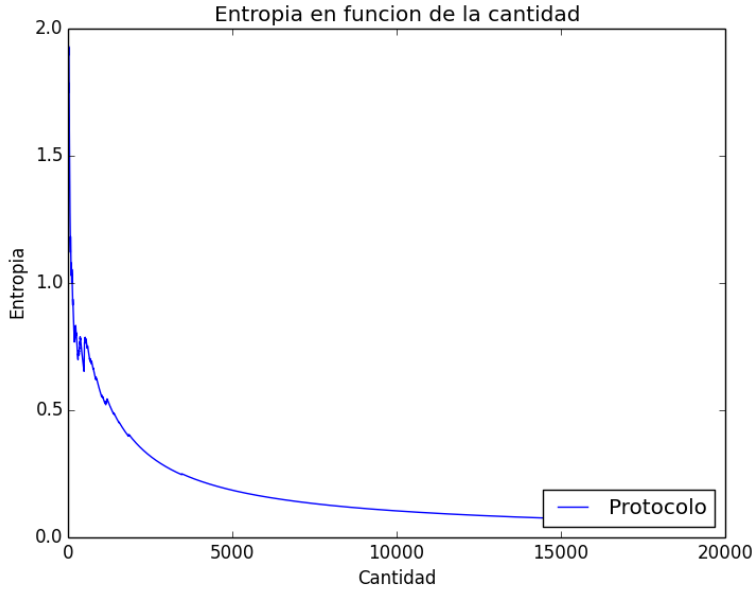
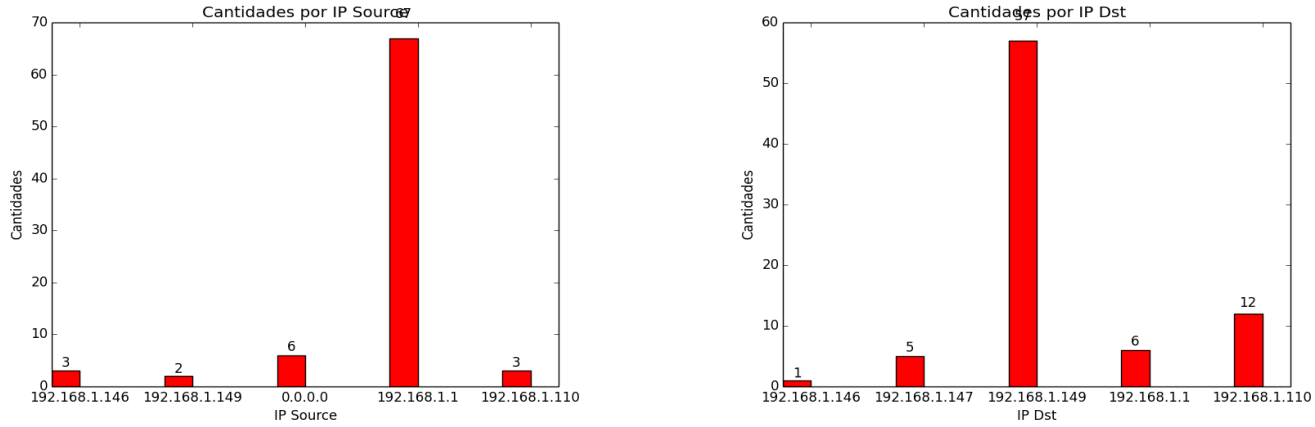


Figura 3: Entropía a medida que aumenta la cantidad de símbolos emitidos por la fuente

Podemos ver que la entropía se relaciona con la cantidad y con la probabilidad pues sabemos por la teoría de la información de Shannon que cuando una fuente emite con mucha probabilidad un símbolo, se vuelve poco informativa, dado que lo que ésta puede emitir se vuelve muy predecible. Lo mismo aplica en este caso para los paquetes IPv4.

Por otro lado vemos que la entropía comienza siendo muy alta, lo que puede indicar que en un principio habría una proporción similar de probabilidades entre los protocolos, pero a la larga el protocolo IPv4 domina la red y la fuente se vuelve altamente predecible, por lo cual la entropía decrece fuertemente.

Veamos a continuación lo que sucedió con los envíos de paquetes IP desde y hacia los hosts.

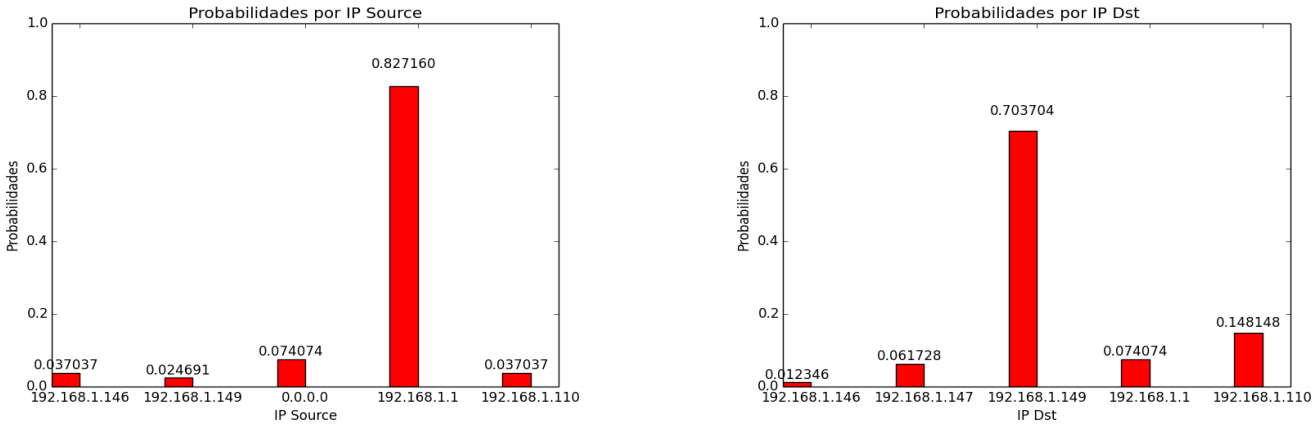


(a) (b)

Figura 4: Direcciones IP asignadas en la red

En esta figura podemos ver que la mayor cantidad del tráfico se lo llevan las direcciones 192.168.1.1 y 192.168.1.149, en donde claramente se ve que interactúan fuertemente entre ellas. Creemos que esto representa el tráfico entre la notebook y el router, para acceder a contenidos de internet. Los demás equipos presentan un rol más pasivo y tiene correspondencia con el hecho de ser celulares o tablets en un ambiente hogareño.

Visto de otro modo, las probabilidades de estos mensajes resultaron las siguientes.



(a) (b)

Figura 5: Probabilidades correspondientes a los paquetes IP origen y destino

Comparando los gráficos de origen y destino podemos ver que los demás equipos efectivamente tienen un rol pasivo ya que se les envía bastantes más paquetes de los que ellos mandan.

Veamos la entropía de la red.

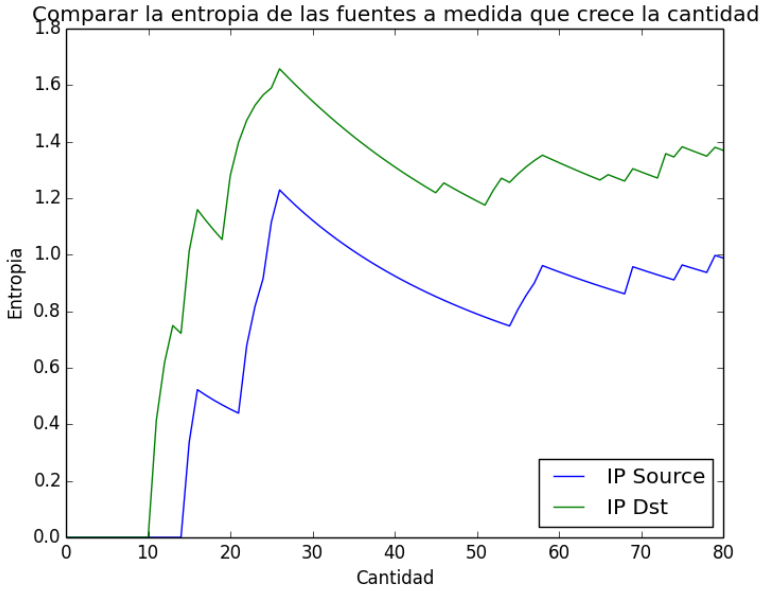


Figura 6: Entropía a medida que aumenta la cantidad de símbolos emitidos por la fuente

El gráfico presenta un comportamiento particular que no observamos anteriormente. En principio creemos que comienza con un valor de 0 y se estanca por un tiempo ya que ningún dispositivo tuvo actividad en la red en ese intervalo (lo cual es entendible sabiendo que hay pocos dispositivos). En cuanto se detecta actividad sube fuertemente la entropía y desde ahí comienza a bajar y subir con un patrón similar. La explicación que encontramos a esto es que en esos picos es cuando se pelean las direcciones 192.168.1.1 y 192.168.1.149 en mandarse mensajes entre ellas. Cuando se envían la misma cantidad de mensajes la entropía sube (ya que crece la incertidumbre), luego una de ellas emite más que la otra y la entropía baja, luego se vuelven a nivelar y así sucesivamente.

El mapa de la red quedó representado de la siguiente forma.

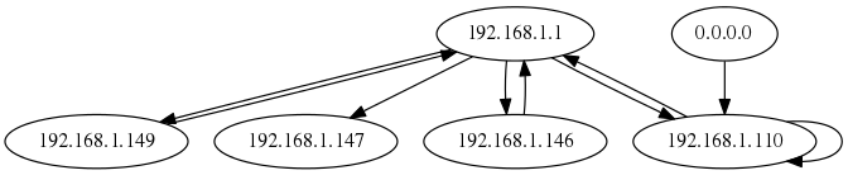
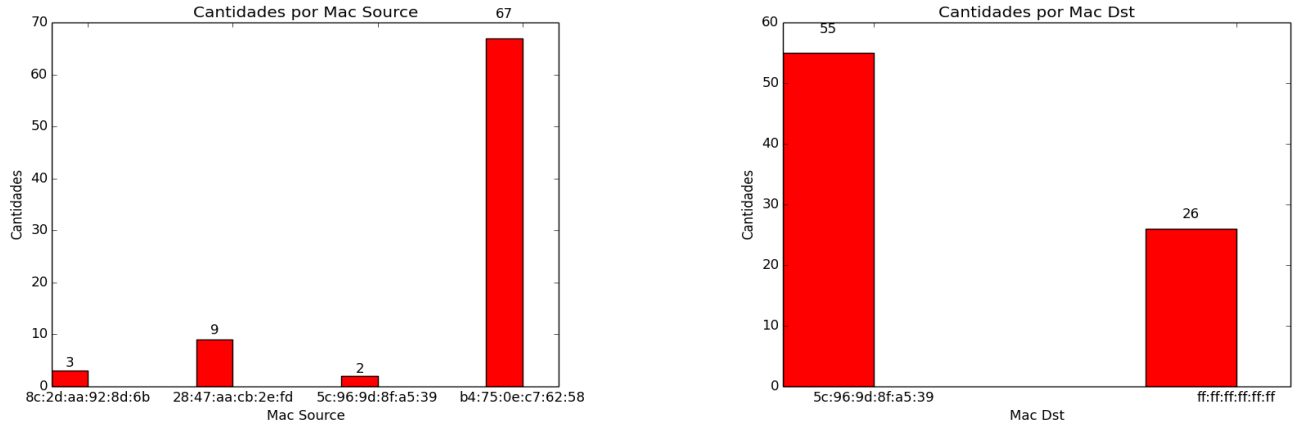


Figura 7: Grafo de conexiones (por envío de mensajes) de la red hogareña

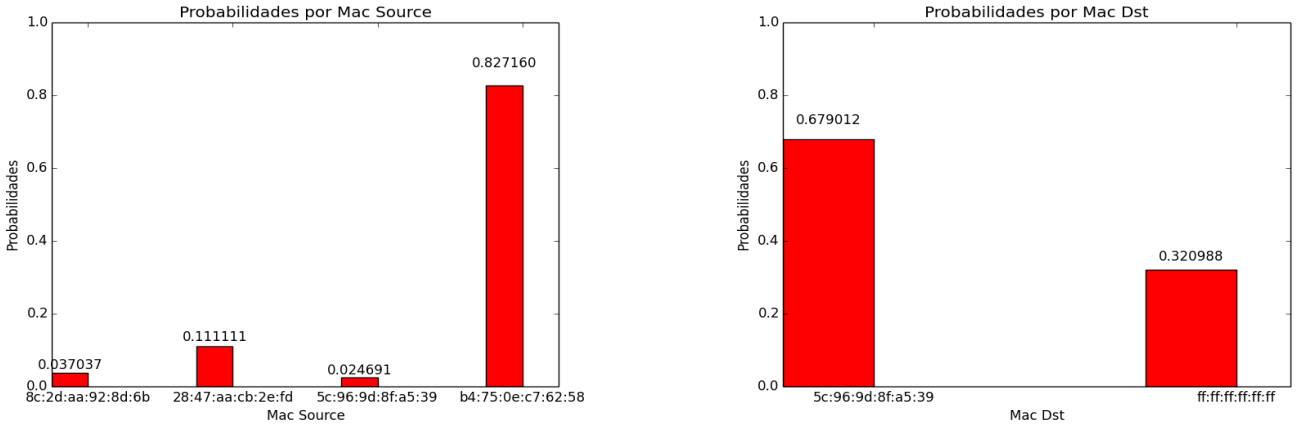
Por último veamos cómo quedó el envío distribuído por direcciones MAC.



(a) (b)

Figura 8: Direcciones MAC asignadas en la red

En este gráfico aparecen menos hosts que en el caso de paquetes IP. Esto, creemos, es porque el router reasigna direcciones IP distintas a un mismo equipo cada vez que se conecta, por lo cual la dirección MAC que envía es siempre la misma.



(a) (b)

Figura 9: Probabilidad por cada dirección MAC de la red de aparecer en un mensaje

Por último veamos la entropía de la fuente.

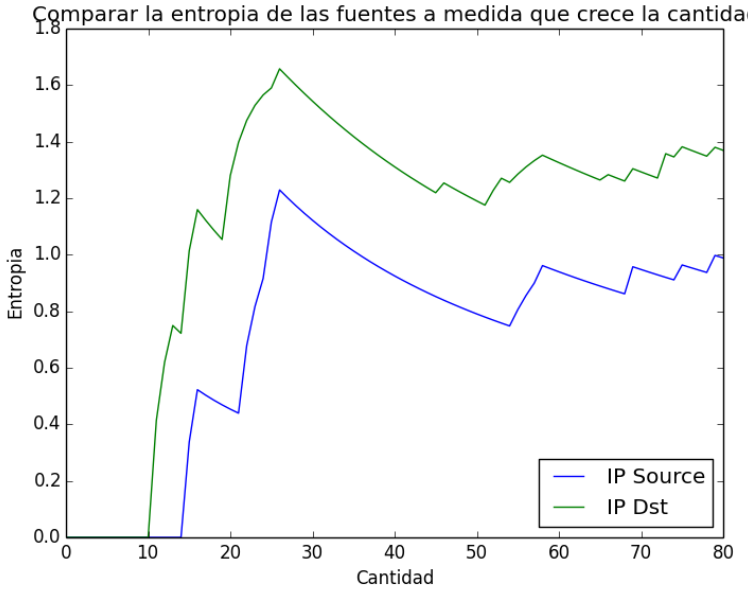


Figura 10: Entropía a medida que aumenta la cantidad de símbolos emitidos por la fuente

Este gráfico presenta una tendencia similar a la de la entropía por dirección IP, con una curvatura en el pico de la MAC destino que antes se presentaba entre los 25 y 30 paquetes. Este fenómeno creemos que puede deberse a dos dispositivos que en ese momento cambiaron a la misma IP que presentó el pico (porque alguno se desconectó o reconectó, por ejemplo) y sin embargo en las direcciones MAC fueron distintas. Por eso la curva de la entropía por MAC no tuvo un pico tan brusco.

Para concluir con esta primera experimentación, podemos rescatar que en una red pequeña de hogar queda corroborado que suele haber poca actividad y la mayor cantidad de paquetes son los IPv4.

3.2. Red del laboratorio de computación

Veamos en primer lugar datos sobre la cantidad y diversidad de protocolos de esta red. Queremos remarcar en primer lugar que, al haber muchos equipos, a modo de mostrar gráficos más legibles, dejamos en vista sólo las direcciones IP y MAC que cumplieron haber mandado al menos 100 mensajes (de lo contrario terminaban siendo al rededor de 80 equipos distintos, muchos con un tráfico insignificante, y los resultados se volvían ilegibles). Además, en los gráficos de probabilidad mostramos los equipos con al menos 0.02 % de los mensajes de la red.

Comenzando con la distinción por protocolos, sabemos que esta es una red que interactúa con muchos dispositivos distintos y por lo tanto podría presentar más protocolos de lo normal. Veamos qué es lo que arrojaron los resultados.

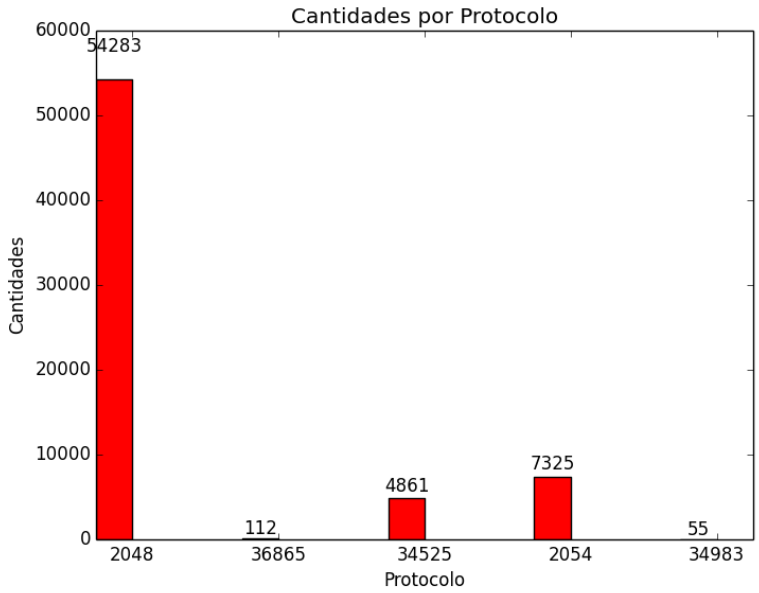


Figura 11: Distintos tipos de protocolos (EtherType) encontrados en la red, organizados por cantidad

Lo que muestra la figura es que, al igual que en la red hogareña, aparecen muy pocos protocolos. Esta vez, en cambio, la proporción de mensajes IPv4 (EtherType 2048) no llega a ser tan desmedida con mensajes de protocolo ARP (EtherType 2054) y de IPv6 (EtherType 34525). Por lo cual la proporción del protocolo ARP sobre la red es del 0,0729 % (comparado con el 0,0021 % del experimento anterior).

Esto le da credibilidad a la hipótesis que planteamos de que la gente en este ambiente suele conectar y desconectar mucho los aparatos de la red, y suele usarla durante pequeños lapsos.

Veamos la misma figura mostrando los respectivos porcentajes.

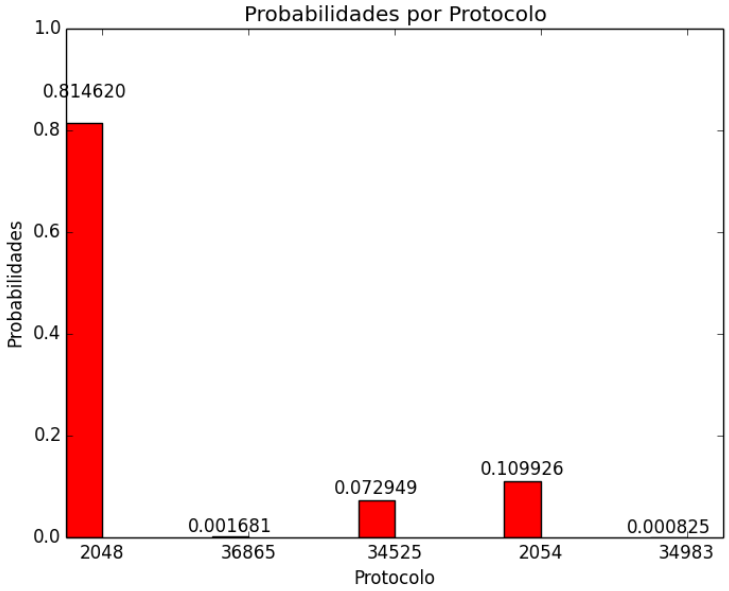


Figura 12: Probabilidad por cada tipo de protocolo detectado

Esta baja respecto del porcentaje de mensajes de protocolo IP nos hace notar la idea hablada en clase de que a medida que aumentan las redes (es decir, aumenta la cantida de equipos de una red), comienza a haber más tráfico de paquetes para sincronizar y organizar y menos de paquetes con información de verdad. Lo cual nos hace comprender por qué es que conviene limitar la cantidad de computadoras en una misma red.

Veamos qué sucedió con la entropía.

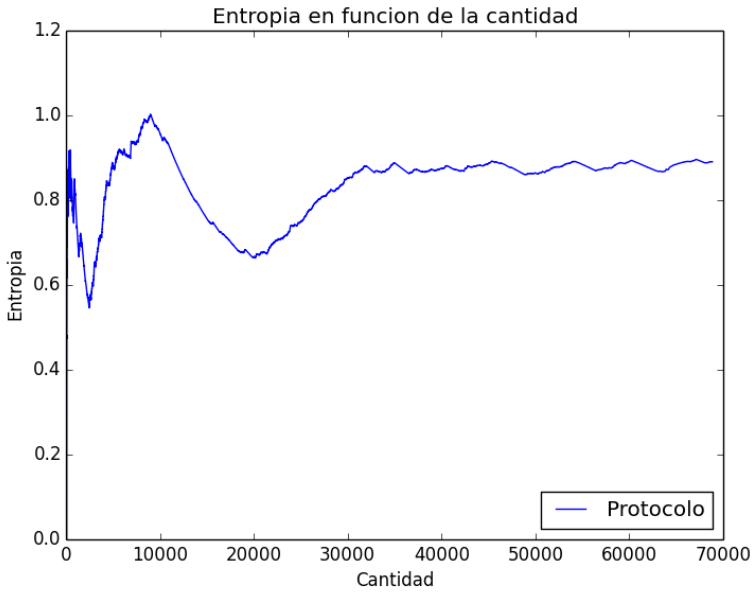


Figura 13: Entropía a medida que aumenta la cantidad de símbolos emitidos por la fuente

Si bien se puede ver que oscila, la tendencia parece ser de estancarse en un mismo valor. Es decir, no parece haber mayores picos ni competencia de popularidad (por así decirlo) por parte de varios equipos. Sí notamos que la entropía tiene un valor bajo y esto podría indicar que hay un equipo (o muy pocos)

que siempre envían mensajes.

Veamos esto con los gráficos de los envíos de paquetes IP desde y hacia los hosts.

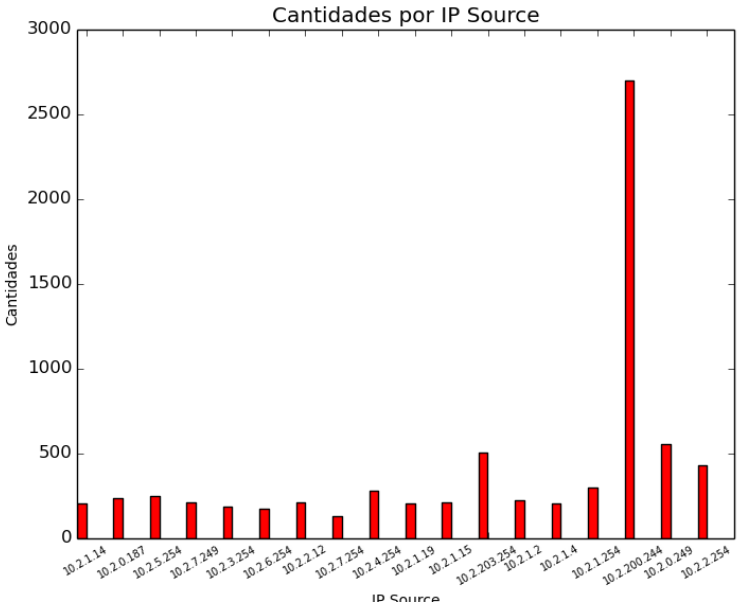


Figura 14: Mensajes con IP origen en la red

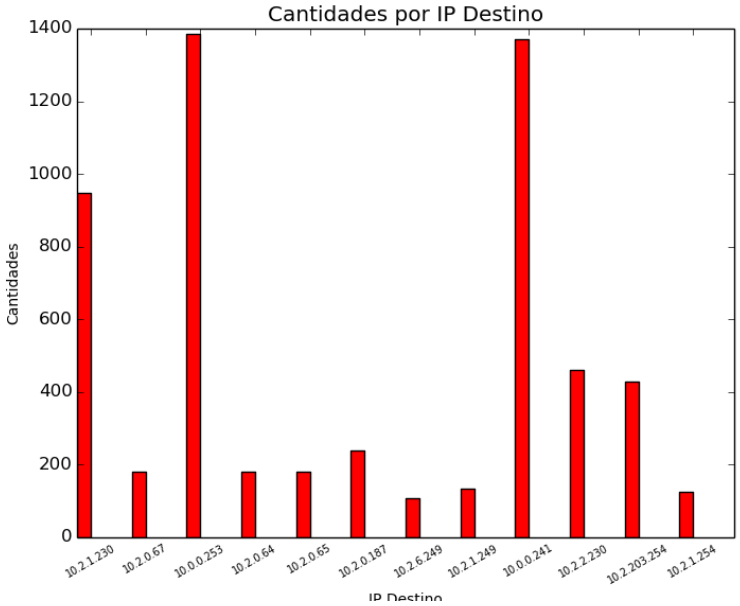
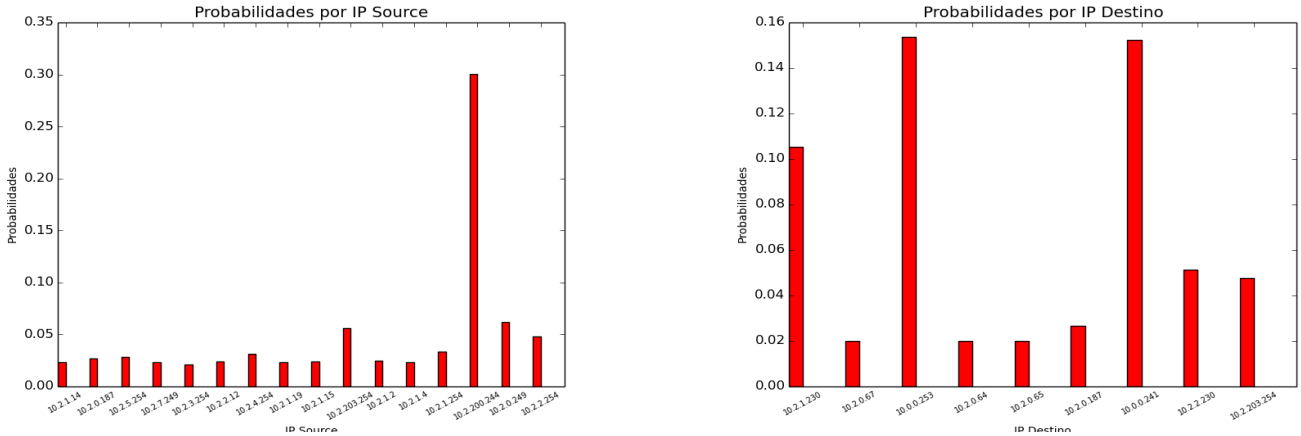


Figura 15: Mensajes con IP destino en la red

En esta figura podemos ver que la mayor cantidad del tráfico de envío se lo lleva la dirección 10.2.1.254, con una fuerte desigualdad respecto al resto de dispositivos. Esto podría representar a un usuario que utilizó una notebook y que estuvo navegando en internet la mayor parte del tiempo en que duró el experimento. Por otro lado el tráfico de recepción estuvo apenas un poco más compartido, por 3 equipos que se distinguieron (10.2.1.230, 10.0.0.253 y 10.0.0.241, entre los cuales no aparece el host anteriormente mencionado).

Visto de otro modo, las probabilidades de estos mensajes resultaron las siguientes.



(a) (b)

Figura 16: Probabilidades correspondientes a los paquetes IP origen y destino

Comparando los porcentajes se ve que en general el uso que le da cada host a la red suele ser equitativo, exceptuando algunos outliers que acaparan la red en su totalidad. Esta desigualdad tan grande con el o los outliers puede ocasionar que la fuente igualmente pierda incertidumbre y por lo tanto la entropía sea baja. Veamos justamente cómo se comportó la entropía en el siguiente gráfico.

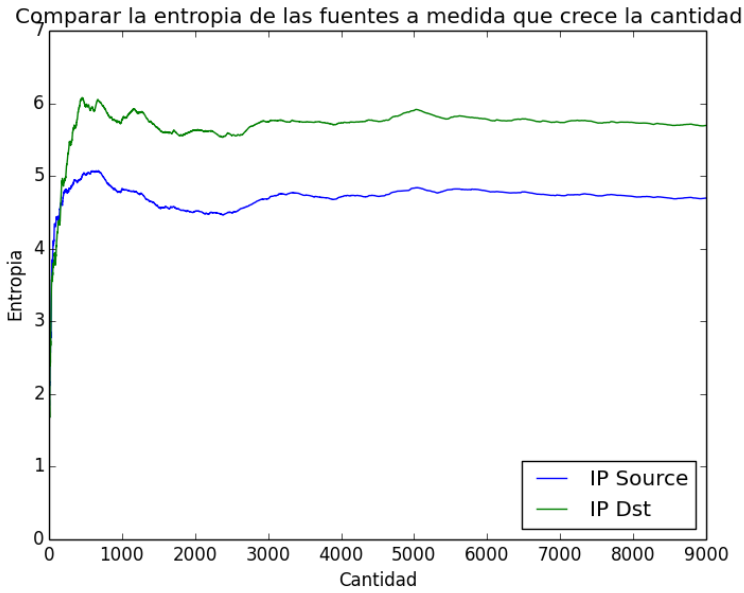


Figura 17: Entropía a medida que aumenta la cantidad de símbolos emitidos por la fuente

Contradiciendo lo que pensábamos en un principio, la entropía resultó tener un valor elevado. Esto creemos que se debe a que la suma de todos los hosts termina consiguiendo un porcentaje bastante superior al del outlier (donde este último tuvo aproximadamente 0.31 % en el caso de IP origen). Por otro lado la entropía siguió estable a través de todo el experimento por lo cual no hubo momentos en los cuales un equipo acaparó totalmente la red. Eso habla bien del nivel de fairness de uso de la red.

Por último veamos qué sucedió desde el punto de vista de las direcciones MAC.

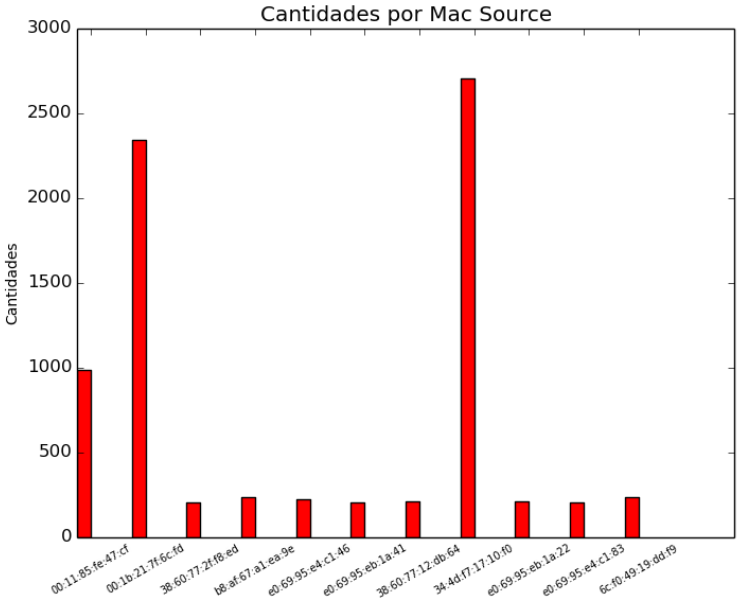


Figura 18: Mensajes de direcciones MAC origen en la red

En este caso aparecieron menos cantidad de dispositivos a lo largo de toda la hora en la que se utilizó la herramienta. Además y por lo que se puede ver en el gráfico, aparecieron más de una dirección dominante (el outlier del que hablábamos antes) y esto nos muestra que dos direcciones MAC muy activas tomaron la misma dirección IP al desconectarse y conectarse. Es por esta suma que se originó el gran outlier antes mencionado. Por otro lado, el tercer dispositivo con mayor cantidad en esta figura muestra un envío de 1000 paquetes, cuando anteriormente vimos que ninguna dirección IP (que no fuera la distinguida) enviaba más de 200 paquetes. Esto explica que ese host se conectó y desconectó al menos unas 5 veces en medio de todo el intervalo de tiempo.

Por último veamos la entropía de la fuente.

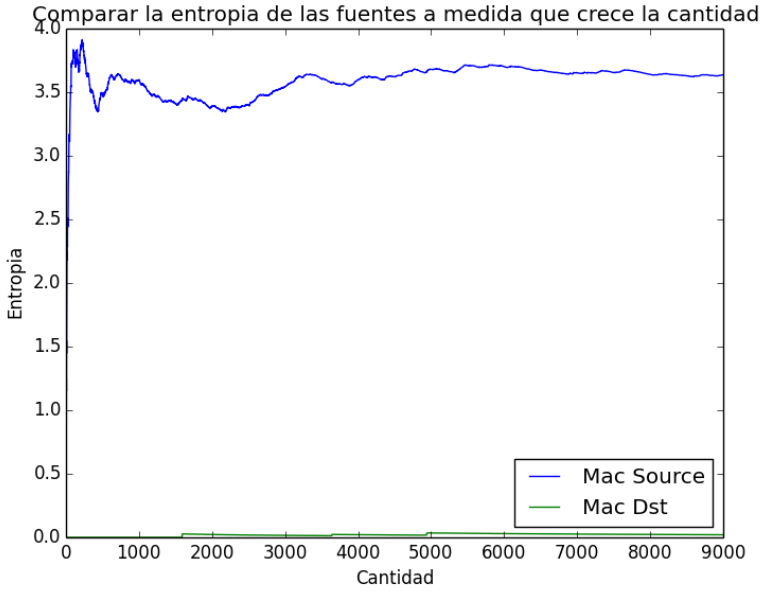


Figura 19: Entropía a medida que aumenta la cantidad de símbolos emitidos por la fuente

Este gráfico muestra la misma tendencia que hubo con las direcciones IP: un valor alto dado que hay incertidumbre por parte de la fuente, y un valor que se estanca en el tiempo, indicando que no hay un momento específico en que un usuario pasa de ser inactivo a dominar la red.

Para concluir esta segunda parte, podemos rescatar varias ideas:

- Durante el mediodía, los laboratorios tienen bastante actividad, aunque no es la actividad que se presenta por la tarde-noche (después de las 17hs). Notamos que a lo largo de la hora en la que ejecutamos la herramienta, se conectaron más de 80 equipos distintos y eso equivale a dos laboratorios completamente ocupados. En comparación, por la noche suelen llenarse los 6 laboratorios.
- Suele haber mucho más tráfico de paquetes ARP y de otros protocolos que en una red hogareña habitual, por lo cual tiende a saturarse por los propios mensajes que impone la misma organización de máquinas en la red.

3.3. Red —

4. Conclusiones

5. Trabajo futuro

A lo largo de este trabajo se nos ocurrieron algunas ideas que creemos sería de gran aporte profundizarlas. Entre ellas están:

- Utilizar la herramienta en algún lugar público como ser un shopping, una cafetería o una convención de tecnología. Nos parece que juntar esos datos con los que obtuvimos nos daría una mejor noción de cómo se comportan las redes en general, en la práctica, a grandes rasgos.
- Con motivos de profundizar el entendimiento del uso de la red de los laboratorios de la facultad, nos parece interesante realizar una nueva prueba en horarios de alta concurrencia (a partir de las 17hs). De esta forma podríamos realizar una buena comparación entre los horarios que se creen menos concurridos y lo más concurridos. También se podría aproximar qué porcentaje de la gente que cursa se junta a hacer trabajos prácticos o resolver prácticas de materias.