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

Teoría de las Comunicaciones

Primer Cuatrimestre de 2015

Trabajo práctico

Wiretapping

Fecha de entrega: 29 de Abril

Integrante	LU	Correo electrónico
De Sousa Bispo, Germán	358/12	german_nba11@hotmail.com
De Sousa Bispo, Mariano	389/08	marian_sabianaa@hotmail.com
Raffo, Diego	423/08	draffo@dc.uba.ar
Tilli, Leonardo	511/00	leotilli@gmail.com

Tabla de contenidos

Introducción

En el siguiente trabajo práctico, se analizará el tráfico de paquetes redes locales. Daremos una visión empírica a los conceptos teóricos de la Teoría de la Información.

Para ello, comenzamos utilizando la herramienta sugerida por la cátedra: *Wireshark*. A través de la misma pudimos corroborar el formato de los paquetes explicados en clase, tomando, o *"sniffeando",* los paquetes de una red doméstica. Luego, generamos nuestra propia herramienta, capaz de generar dos tipos de fuentes, así como sus estadísticas asociadas.

Contexto teórico

La entropía mide la cantidad de información media que emite una fuente. El máximo valor de la misma se alcanza cuando todos los símbolos de la fuente son equiprobables. Puede interpretarse también como una medida de incertidumbre, a mayor entropía mayor incertidumbre.

Para calcular la entropía se utiliza la ecuación de Shannon:

$$H(X) = \sum_{i \in X} P(S_i) I(S_i)$$

Donde $P(S_i)$ es la frecuencia muestral del símbolo $i \in S$ e $I(S_i)$ es la cantidad de información del símbolo i. Estos valores se definen de la siguiente manera:

$$P(S_i) = \frac{\#S_i}{\sum_j \#S_j}$$

$$I(S_i) = -log_2(P(S_i))$$

Fuentes de información

De los datos recopilados, se generaron dos fuentes de información. La primera consiste en obtener los campos *type* de cada paquete *ethernet*. La segunda, fue elegida por el grupo.

ARP (o Address Resolution Protocol)[1] es un protocolo para mapear direcciones IP a direcciones físicas (conocidas como Media Access Control o MAC address) reconocibles en las redes locales. Es decir, asocia una dirección de capa 3 (red) con una dirección de capa de enlace o capa 2.

Nos encontramos ante el problema de decidir qué fuente de información obtenida del ARP será útil para analizar cada red. El objetivo es: poder distinguir los nodos, su importancia y accionar que tienen dentro de ella.

Es por esto que decidimos conveniente utilizar como parte de nuestra fuente las siguientes partes del paquete ARP:

-Opcode
-Source Address

Además, utilizamos el *target address* para poder generar el grafo de relaciones.

-Opcode: Los dos valores que suele tener este campo son 1 y 2. Estos corresponden a *request* y *reply* respectivamente. Se utilizan para indicar que se está realizando un pedido de información o una contestación del mismo. De esta manera, podremos saber en qué sentido se realiza la conexión entre los nodos de la red. A priori, creemos que el nodo que más *request* reciba (más *replies* efectúe), será el considerado *access point*. Se validará esta hipótesis con el análisis empírico.

-Source address: Nos permite saber de dónde proviene el paquete. Conjuntamente con el opcode podremos saber si se ha insertado un nodo nuevo a la red, si hay algún nodo especial que realice muchas preguntas o si alguien está refrescando la información qué posee de la red.

-Target address: Nos permite saber a quién va dirigido el paquete. Sin esta información, no podemos plantear un grafo donde las aristas representen la comunicación entre dos nodos cualquiera de la red. Si bien esta información no fue utilizada para el análisis de la fuente, sí fue utilizada para el grafo.

[1] https://msdn.microsoft.com/es-es/library/cc758357%28v=ws.10%29.aspx

Mediciones

Tanto para la fuente de información *S* brindada por la cátedra y la creada por el grupo (para cada muestra realizada), se calculó la frecuencia muestral de cada mensaje, la entropía máxima, y la entropía de la fuente.

Se midieron cuatro redes distintas: una red doméstica (denominada *Local Network* durante las mediciones), una red laboral (*Work Network*), la red de un café (CoffeeHouse Network) y una red de un centro comercial (ShoppingMall Network).

Los paquetes de la red laboral se obtuvieron a partir de escuchar la red LAN de la oficina, mientras que las demás muestras se obtuvieron escuchando redes Wi-Fi.

Esperamos a priori, al analizar cada fuente, que la entropía diste de la máxima. Mientras menos cantidad de nodos distinguidos haya, menor será la entropía: Si un nodo es distinguido, implica que aparece muchas veces en la fuente. Por lo tanto, el grado de incertidumbre ante la aparición de un símbolo, disminuye (es más probable que aparezca el nodo distinguido). Cuanto menor es la entropía, más fácil es predecir qué símbolo aparecerá.

A continuación se detallan los resultados obtenidos para las redes. Para entender la topología de la red, realizamos grafos de relaciones entre sus nodos, discriminados por IP. Se agregan gráficos de torta para analizar por porcentaje, el tipo de paquete en la red. Además, gráficos de barra para la información de la tupla <OpCode, Source>, contrastando con entropía de la fuente, y entropía máxima.

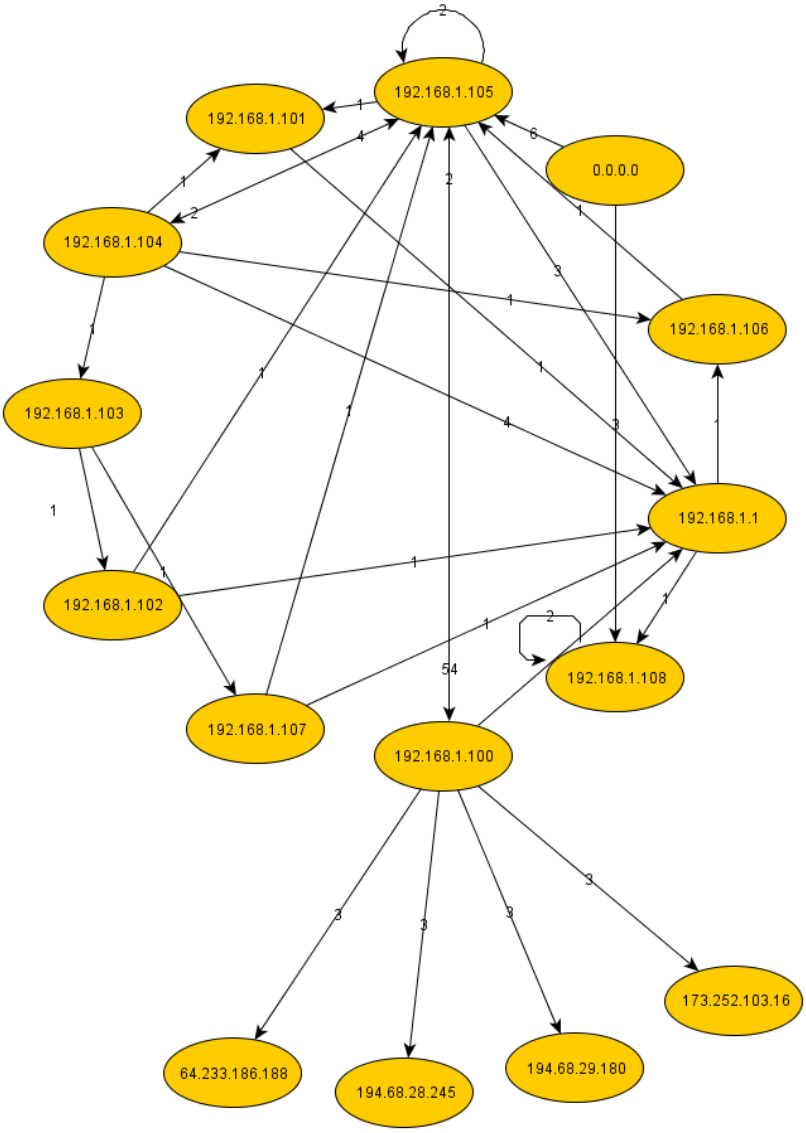
Analizaremos primero por tipo de gráfico, para luego, generar una conclusión general.

NOTA: El cálculo de la entropía fue calculado para todos los nodos analizados. En el caso del análisis de ARP, por gráfico y por declaratividad, se discriminaron por *opcode*, no así el cálculo de la entropía.

Gráfos: Topología de la red

A continuación describiremos la topología de las redes de las redes intervenidas. Para esto se crearon grafos dirigidos indicando los envíos de paquetes *request* entre los distintos nodos de la red.

Local Network



Recordemos que el objetivo del protocolo ARP es encontrar la MAC Address de una computadora dentro de una red local. Los paquetes MAC deben morir dentro de una red LAN/WLAN. De este gráfico distinguimos las siguientes características significativas:

- Forma circular que generan las relaciones de las IP's en la red 192.168.1.0/24.
- Muchos nodos tienden a buscar al dispositivo con IP 192.169.1.105 y 192.168.1.1: La 105 resulta ser una computadora, y la 1 el *access point* de la casa. Por lo tanto, la cantidad de aristas entrantes, no definen necesariamente al router de la red.
- Hay nodos con ciclos: Por la información recopilada en la página de WireShark[2] se utiliza para hacer lo que se llama: *Gratuitous ARP*. Cabe destacar que no se encuentra en la especificación del protocolo ARP. Su objetivo es detectar colisiones de direcciones IP. Una máquina envía un mensaje ARP a sí misma, en caso de que alguien le responda, implica que existe otra máquina con la misma dirección IP, por lo tanto, existe un conflicto.
- Existen pedidos de ARP para direcciones IP fuera de la red. Si la red se encuentra correctamente configurada, esto no debería suceder. El dispositivo 192.168.1.100 corresponde a un teléfono celular: suponemos que durante un pequeño lapso, la máscara de red fue 0.0.0.0, por lo tanto, al enviar un paquete IP, el destinatario matcheaba, aún cuando pertenecía efectivamente a la red. Dado que no existe respuesta del ARP enviado, podemos asegurar que esos dispositivos nunca recibiendo ese mensaje.
- EL envío de ARP entre 192.168.1.105 y 192.168.1.100 es sustancialmente más alto que el resto. Investigamos en internet y encontramos este post[3], que si bien no sirve de fuente confiable, vale la pena mencionar: Dependiendo del fabricante del software, la cantidad de ARP enviados puede variar. Investigando el router de nuestra red local, vemos que la periodicidad con la que se actualizan los ARP varía: por el uso que le estamos dando a la red, parece que para los dispositivos "más estables", se alarga el tiempo de vida de la relación IP-MAC. A continuación se agrega el detalle de la tabla de Active IP al momento del desarrollo de este informe (las IP's no necesariamente coinciden con el grafo, ya que corresponden a días distintos).

DHCP Active IP Table

DHCP Server IP Address: **192.168.1.1** Refresh

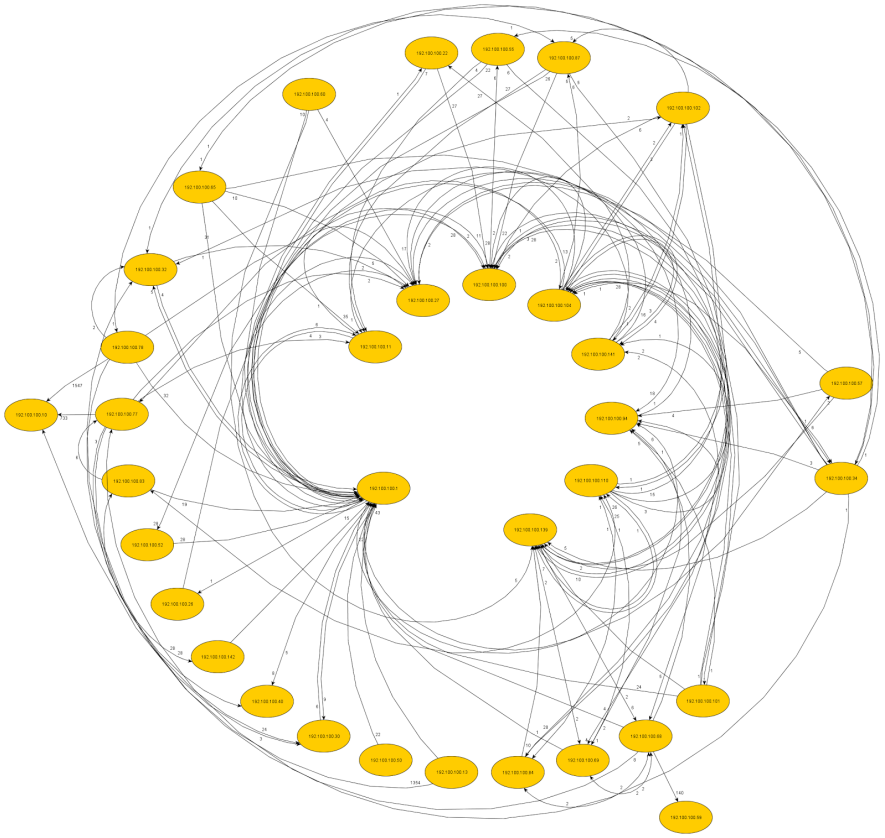
Client Host Name	IP Address	MAC Address	Expires	Delete
emma	192.168.1.105	60:36:dd:98:05:e4	23:09:43	
android-4284de1f0c7fc487	192.168.1.100	1c:af:05:36:07:69	23:08:39	
android-771596ec0a28a827	192.168.1.102	1c:af:05:36:0b:47	23:03:01	
android_f462492e7980808c	192.168.1.108	80:96:b1:c2:22:85	19:08:09	
mtqp_note	192.168.1.104	84:a6:c8:76:77:85	17:41:45	
Enano-ultrabook	192.168.1.106	68:94:23:a5:47:d5	16:01:19	

Close

[2] https://wiki.wireshark.org/Gratuitous_ARP

[3] https://social.technet.microsoft.com/Forums/en-US/388dad56-117c-4f0a-8dc9-6496f5d2b921/getting-continous-arp-requests-from-a-single-pc?forum=netmon

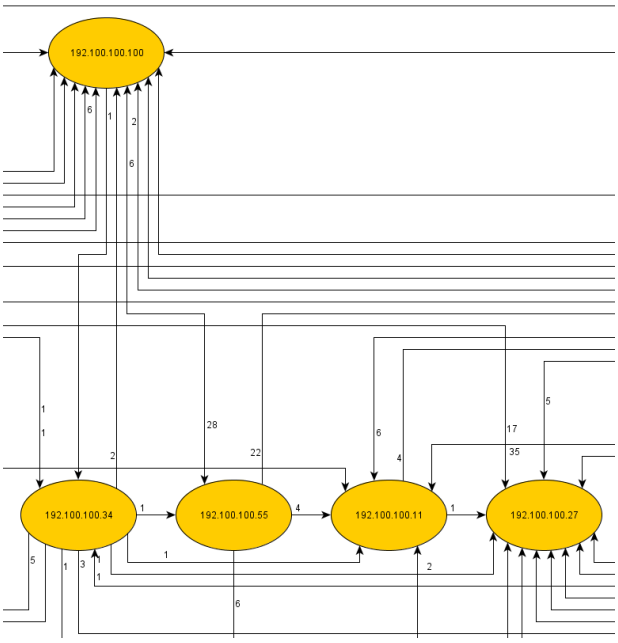
Work Network

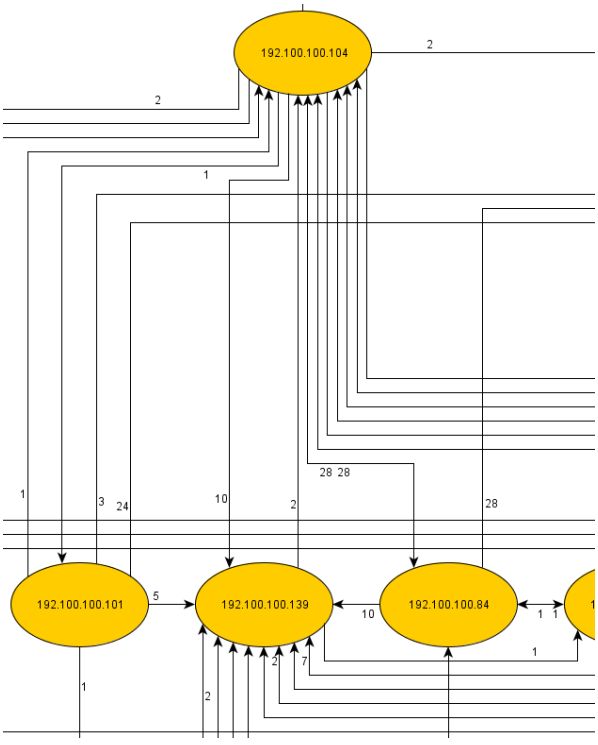


El gráfico anterior es muy grande para que quede claro en el informe, es por eso que se adjunta en la entrega. Al final del informe se encuentra el detalle del entregable para poder inspeccionarlo.

Esta red corresponde a una empresa de desarrollo de software. Posee alrededor de 15 empleados. Queda aún menos claro quién es el *access point*. Conociendo la red, podemos decir que además del *router* a internet, existen varios servidores de base de datos, repositorios de código/especificaciones, servidores FTP, como servidores que hostean entornos de *testing*. Dicho esto, consideramos razonable la aparición de varios nodos significativos en esta topología. Ubicamos estos nodos en el centro, y en el anillo exterior, las computadoras que utilizan estos servidores.

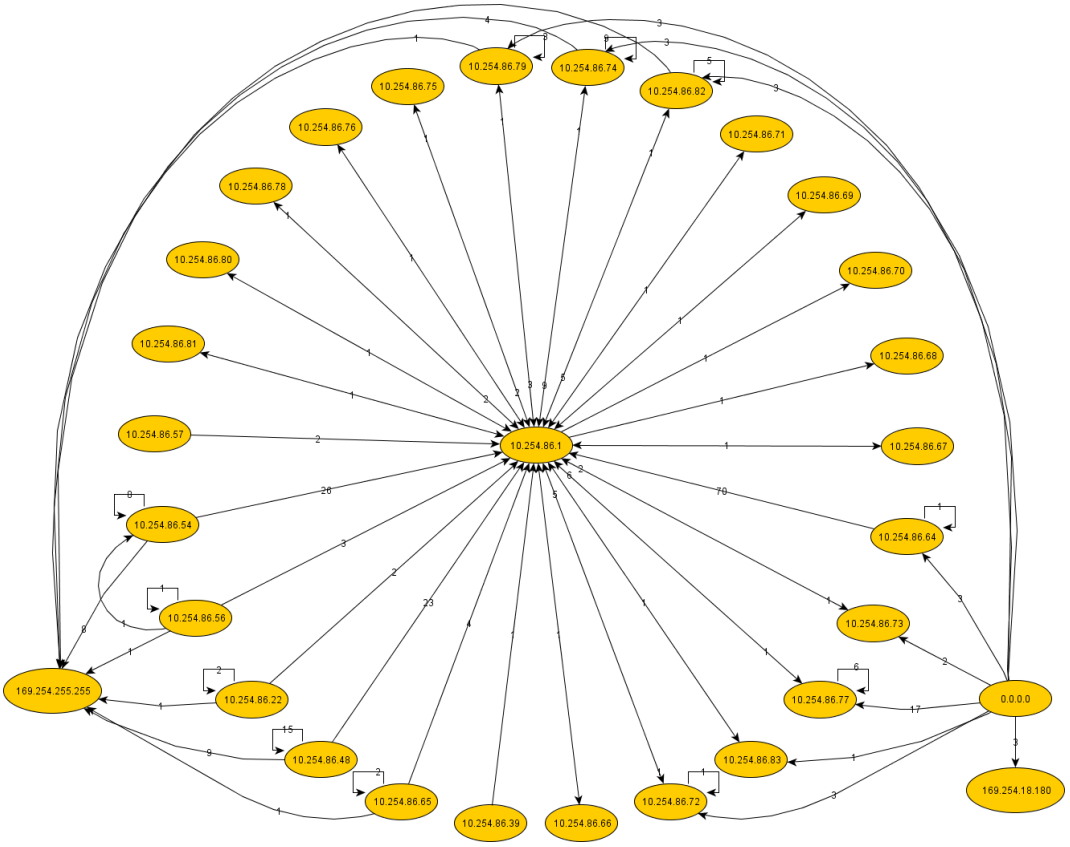
Con el objetivo de clarificar, se reorganizaron los nodos en una estructura jerárquica. El diagrama completo es muy grande, por lo que se adjunta también. Agregamos dos imágenes de dos de los nodos más relevantes de toda la red.





En los gráficos de barra que se detallarán a continuación, estas dos IP’s deberían darnos menor información que por ejemplo, 192.100.100.65, computadora de un desarrollador.

CoffeeHouse Network

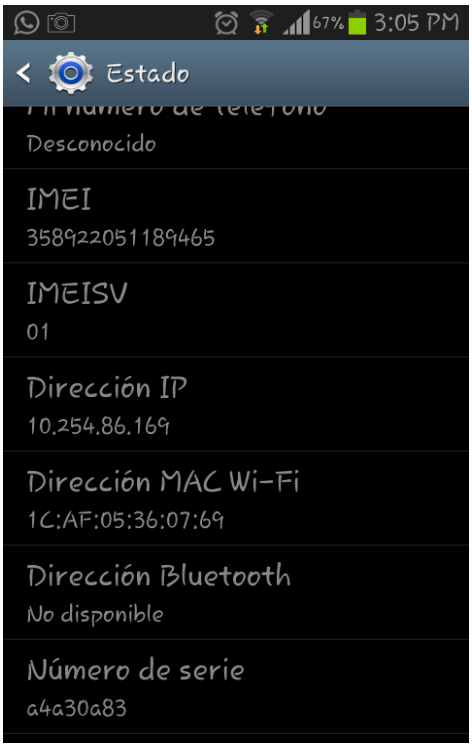


Debido a ser una red de acceso público proveída a través de *FibertelZone*, no tenemos conocimiento a priori de la red. Sin embargo, a partir del grafo armado que la representa podemos ver lo siguiente:

- La IP 10.254.86.1 recibe la mayor cantidad de request de un mismo nodo (70 por parte del nodo con IP 10.254.86.64) y además, es la que aglomera la mayor cantidad de los que parecen ser dispositivos. Claramente es un nodo distinguido y posiblemente un router de la red.
- Podemos encontrar, al igual que en las redes anteriores, que hay nodos que se unen a sí mismos (*Gratuitous ARP*).
- Notemos que existen dos nodos ajenos a la red con valores de IP 169.254.255.255 y 169.254.18.180. Esto se corresponde con *Automatic Private IP Addressing (APIPA)*. Esto implica que si el usuario de la red no puede conseguir su dirección IP utilizando DHCP, el cliente seleccionará una dirección aleatoria entre la IP 169.254.1.0 a 169.254.254.255 (inclusive), con una máscara de red de 255.255.0.0. Luego enviará paquetes ARP preguntando a que dirección de capa 2 corresponde. Si ya existe una, generará otra dirección IP aleatoria.

Todo el rango de direcciones 169.254.0.0/16 fueron reservadas y son denominadas direcciones “link local”. La presencia de estas direcciones, en muchos casos, implican la pérdida de conectividad a la red, o que el servidor DHCP está caído.[4]

Mientras realizabamos las mediciones, averiguamos cuál era la dirección IP de la máquina en la que nos encontrabamos y la misma era: 169.254.189.129 con máscara 255.255.0.0. En el momento creímos que se trataba de nuestra dirección IP. Sin embargo, al encontrarnos con el grafo generado para la red, nos generó dudas la ausencia de esa dirección, por lo que fuimos de vuelta y para ver qué dirección IP se asignaba, encontrando la siguiente:



- Otro nodo que es interesante de analizar es el que posee la IP 0.0.0.0. Una posible explicación para esto es que, al momento de conectarse a la red, el nodo todavía no tiene asignada una dirección IP, por lo que envía un ARP con dirección nula, hasta que se le sea asignada una por el cliente DHCP.[5]

[4] https://wiki.wireshark.org/APIPA

[5] https://ask.wireshark.org/questions/5178/why-gratuitous-arps-for-0000

ShoppingNetwork

A continuación se presentará la tabla con el tráfico de paquetes ARP para los datos recolectados en un centro comercial de la provincia de Buenos Aires. No incluímos el grafo, ya que no consideramos que vaya a agregar información adicional.

OpCode	Source	Destination	Count	Probability	Event Information
1	'172.17.0.3'	'172.17.62.182'	1	0.2	2.32192809489
1	'172.17.62.182'	'172.17.0.3'	1	0.2	2.32192809489
2	'172.17.0.3'	'172.17.62.182'	2	0.4	1.32192809489
2	'172.17.62.182'	'172.17.0.3'	1	0.2	2.32192809489

Esta muestra se tomó durante un lapso de 15 minutos, su objetivo fue analizar el comportamiento de una red potencialmente muy grande, en situaciones cotidianas: dispositivos conectados por lapsos cortos de tiempo. Llama la atención que sólo se hayan encontrado 2 nodos que emitían y respondían paquetes ARP.

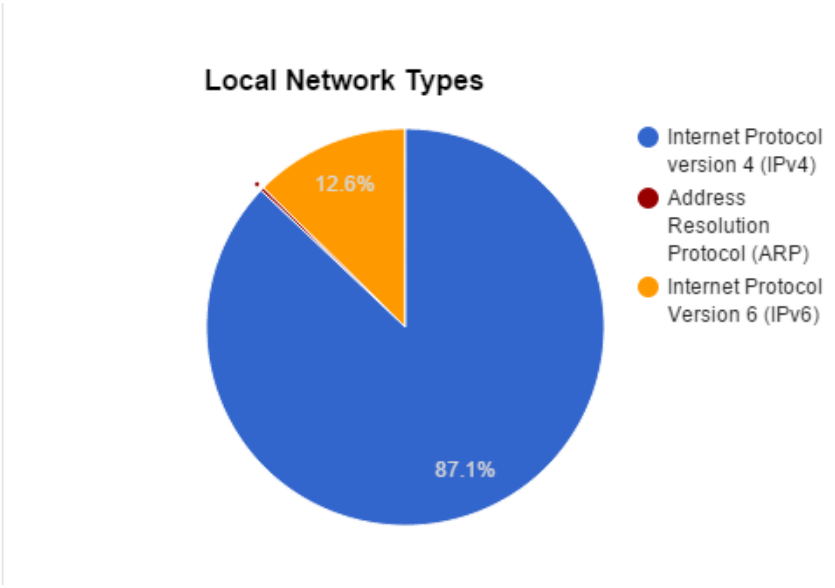
Al igual que la CoffeeHouse Network, esta red contaba con Fibertelzone.

Se analizará esta red más detenidamente en los siguientes incisos.

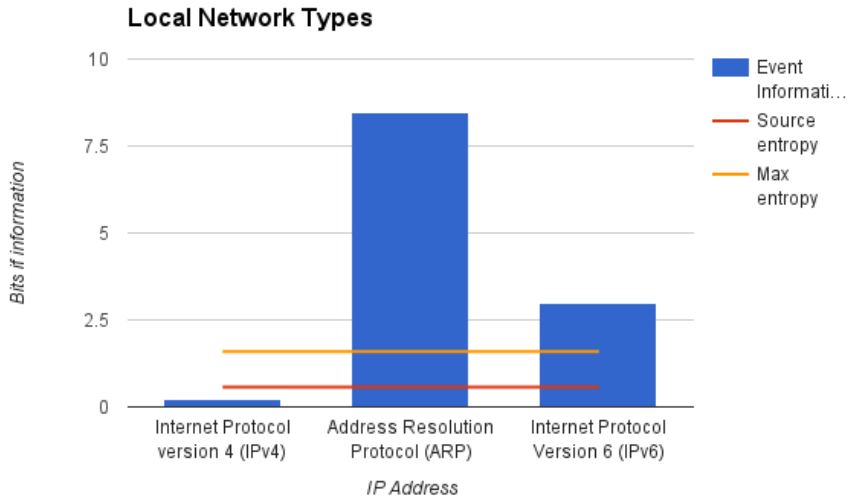
Análisis de tipos de paquetes

A continuación se mostrarán 2 tipos de graficos referidos a los tipos de paquetes que circulaban por cada una de las redes. El primero de los gráficos, mostrará la proporción de cada tipo de paquete, y el segundo, la cantidad bits de información de cada tipo de paquete, comparándola con la entropía de la red.

Local Network



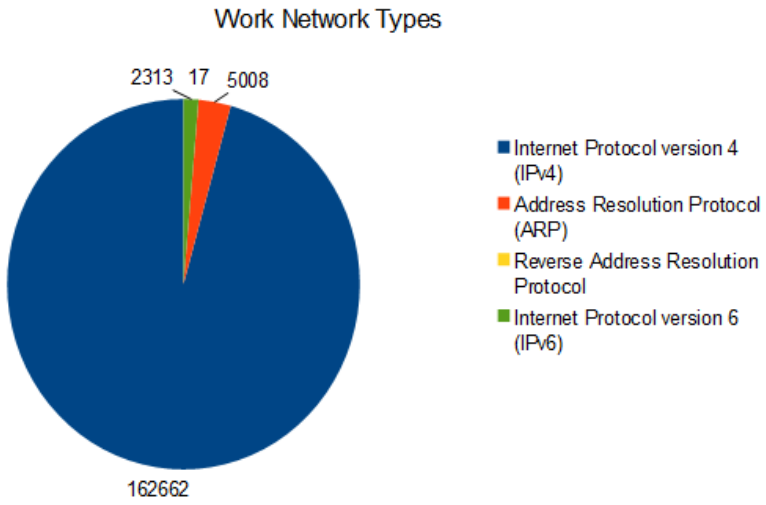
Como se puede observar en el gráfico, la relación entre los tipos de paquetes no es uniforme, sino que el **87.1%** de estos son de tipo IPv4, el **12.6%** son IPv6 y sólo el **0.3%** restantes corresponden a mensajes de tipo ARP. Esto nos habla de la estabilidad de los nodos en una red de pequeño tamaño, donde no hay nuevos nodos incorporándose continuamente, y las comunicaciones entre los nodos no distinguidos no son tan asiduas.



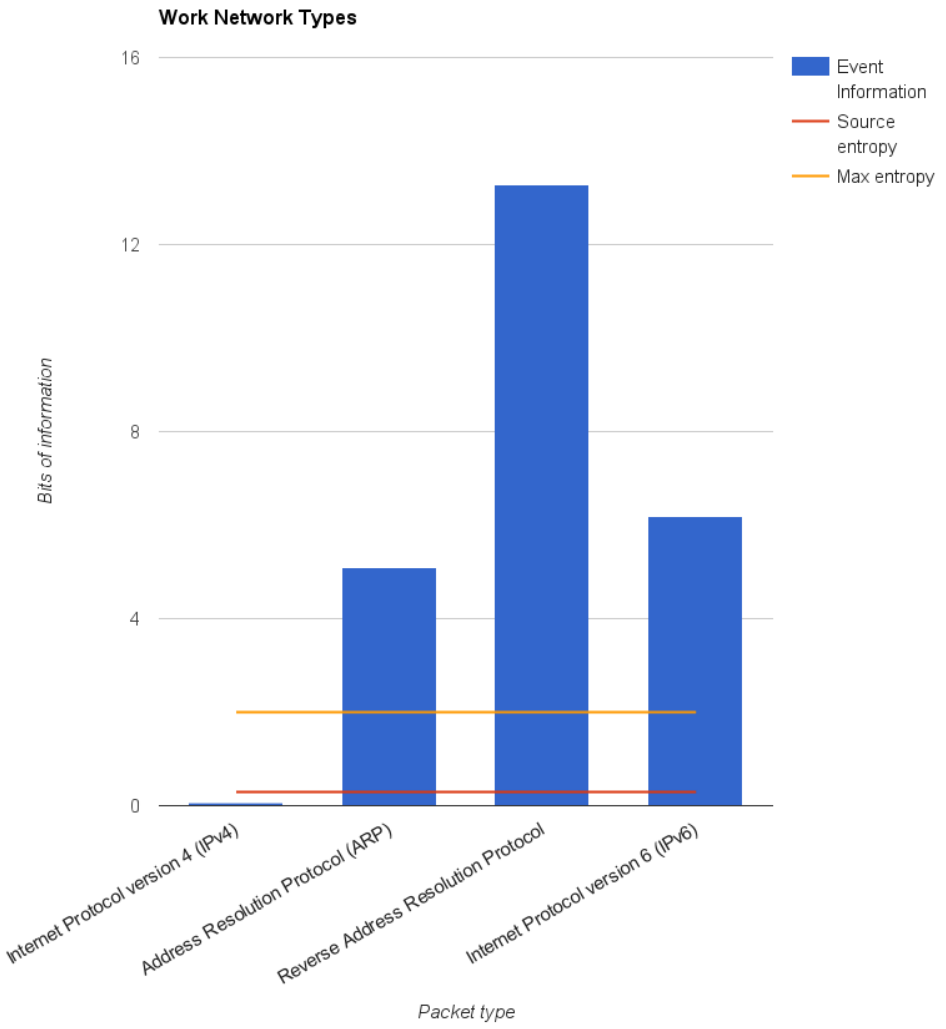
Este gráfico nos muestra la información que brinda cada paquete, contrastado con la entropía de la fuente. En este caso, los bits de información proporcionados por los paquetes ARP son significativamente mayores al resto, y muy por encima de la entropía de la fuente.

Analizando estos dos gráficos podemos apreciar que el Protocolo IPv4 es un protocolo distinguido de la red, y además que ARP, impone muy poco *overhead*.

Work Network



Se puede observar que en la *Work Network* la proporción de los tipos de paquetes no es homogénea, siendo un **95.8%** paquetes de tipo IPv4, 2.9% paquetes de tipo ARP, 1.3% paquetes de tipo IPv6 y 0.01% paquetes de tipo RARP. Sin embargo, si comparamos con los porcentajes de la red hogareña, veremos que la proporción de paquetes ARP es significativamente mayor. Esto puede deberse a que el tráfico interno en la red laboral es mayor a la hogareña, con lo que necesariamente, la actualización de las MAC debe ser más frecuente, y entonces, el porcentaje de ARP es mayor.



En este gráfico podemos ver nuevamente como los paquetes IPv4 aportan menor cantidad de información debido a la desproporcionada probabilidad de aparición de estos. Los paquetes de tipo RARP, que fueron sólo 17 aportan una gran cantidad de información ya que su probabilidad de aparición es extremadamente pequeña, es decir, no son paquetes comunes en el tráfico de esta red.

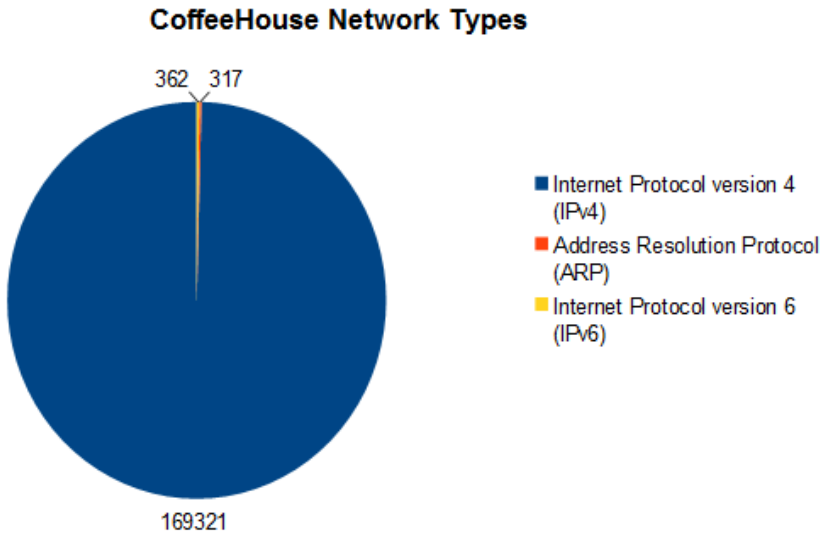
Los paquetes RARP (Reverse Address Resolution Protocol) son aquellos paquetes que se envían para averiguar una IP sabiendo de antemano una MAC Address. Estos paquetes son

comúnmente enviado en el momento de bootear un nodo de la red, ya que sabiendo su MAC Address necesita averiguar su IP Address[6].

Analizando la red, se encontró que los paquetes RARP se envían debido al uso de maquinas virtuales (*VMWare*), que se utilizan frecuentemente en esta red. Esto se puede observar en la siguiente imagen tomada desde WireShark.

No.	Time	Source	Destination	Protocol	Length	Info
156278	2183.480011	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156280	2183.690609	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156284	2184.690566	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156294	2185.690577	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156303	2187.690646	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156328	2190.690701	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156351	2195.690809	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156414	2203.690932	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156483	2210.691047	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156505	2211.691057	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156507	2212.691098	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156553	2214.691079	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0
156573	2217.691102	Vmware_6a:3e:c0	Broadcast	RARP	60	Who is 00:0c:29:6a:3e:c0? Tell 00:0c:29:6a:3e:c0

CoffeeHouse Network

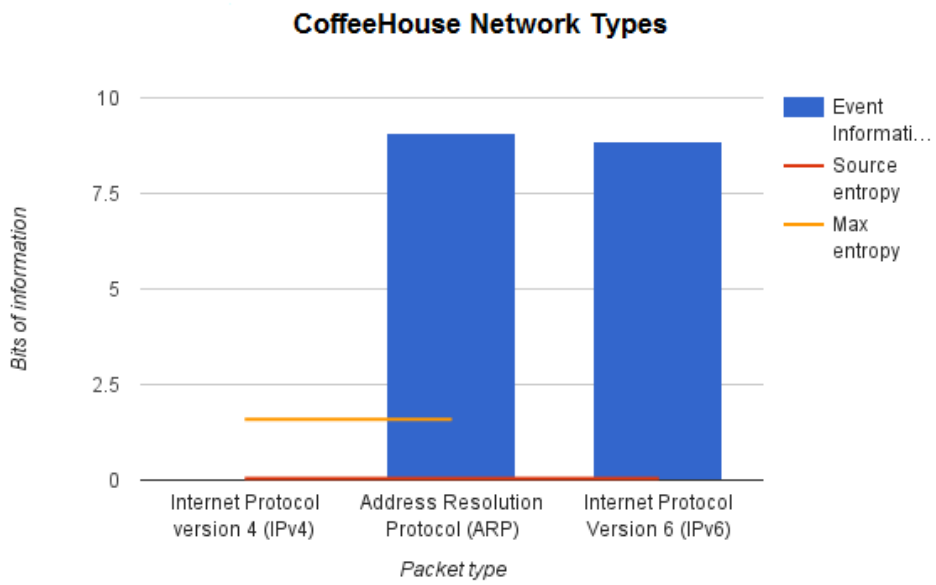


Siguiendo el patrón de las demás redes, podemos ver que la cantidad de paquetes de tipo IPv4 (**99.6%**) obtenidas en la red CoffeeHouse es ampliamente superior a los demás tipos de paquetes, obteniendo un valor parecido (dado el tamaño muestral) para los paquetes ARP(**0,02%**) e IPv6 (**0,02%**).

La mayor comunicación entre los nodos se produce contra el *router* y no entre los nodos no distinguidos. La mayoría de los usuarios de la red simplemente utilizan internet y

[6] http://www.netometer.com/qa/rarp.html#A9

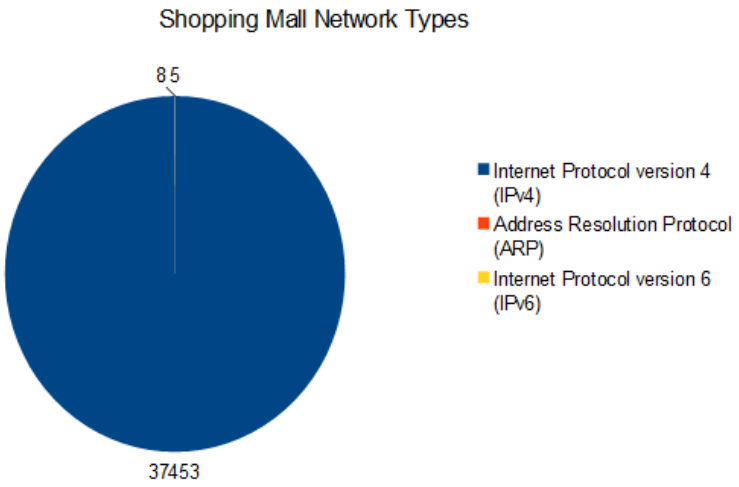
no para realizar comunicaciones entre sí.De esta forma, la cantidad de paquetes ARP no debería ser grande en comparación con la cantidad de paquetes obtenidos. Así mismo, es esperable que la cantidad de paquetes IPv6 no sea alta, por estar en un período de transición desde la tecnología previa.



Como bien se espera, la cantidad de información del tipo ARP e IPv6 es muy alta por lo que la cantidad de paquetes obtenidos del tipo son muy bajos, lo cual vimos en el gráfico anterior. A su vez, como habíamos mencionado, la cantidad de paquetes encontrados para ambos tipos era similar para el tamaño muestral. Esto se refleja en una poca variación del valor de la información.

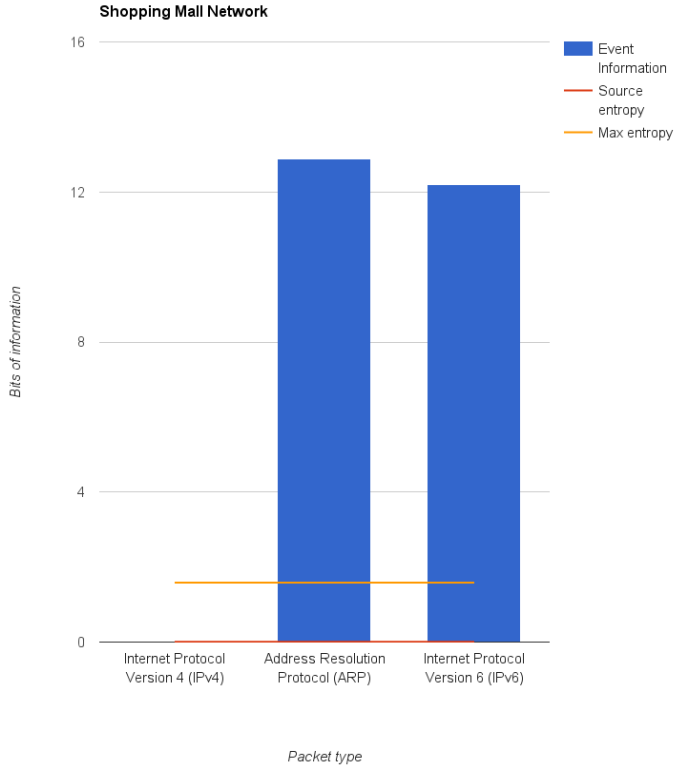
Por el contrario, habíamos visto que la cantidad de paquetes IPv4 era muy superior al resto, ocupando casi el porcentaje total de la muestra. Es por esto que el valor de la información es casi nulo, ya que el grado de incertidumbre del símbolo es muy bajo.

Shopping Mall Network



Podemos ver que esta red exhibe, junto a la CoffeeHouse Newtork, la relación más desproporcionada entre paquetes IPv4 **99.97%,** IPv6 **0.02%** y paquetes ARP **0.01%**.

Esto se puede explicar dado que la medición se hizo por un tiempo muy acotado, y si bien la red era potencialmente grande, en ese tiempo la red se mantuvo estable, no habiéndose sumado nodos. Además, el comportamiento esperado en este tipo de redes es la utilización de internet, y no la comunicación entre dispositivos conectados a la red local.



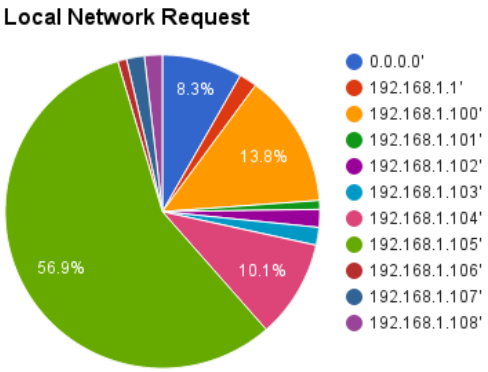
Lógicamente, la cantidad de información que proporciona el protocolo distinguido de esta red, que es IPv4 (**0.0005 bits)** está muy por debajo de la entropía (**0.0048 bits**) , y dista mucho la información proporcionada por un paquete de otro protocolo (cuyos valores rondan los **12 bits**). La entropía de esta red se encuentra en un orden de magnitud superior al de la información del protocolo distinguido de la red.

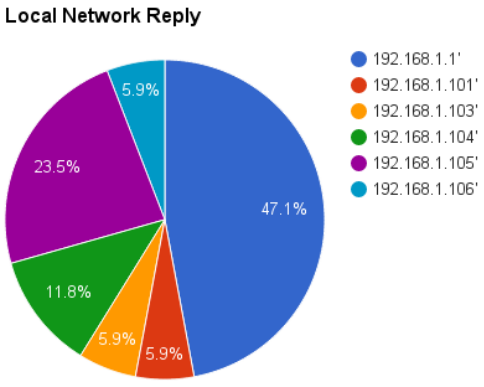
Trafico ARP

Para analizar el tráfico ARP, se generaron gráficos donde se muestra la proporción de mensajes enviados por cada nodo de la red y, por otro lado, la cantidad de bits de información que brindan contrastada con la entropía de la red.

Estas últimas mediciones se hicieron tomando como símbolo la tupla (opCode, SourceAddress), y la entropía fue calculada en consecuencia. Para poder analizar mejor la información, se separó el análisis en paquetes de tipo ARP *request* y paquetes de tipo ARP *reply*, pero la entropía con la que se contrasta es la misma en ambos casos.

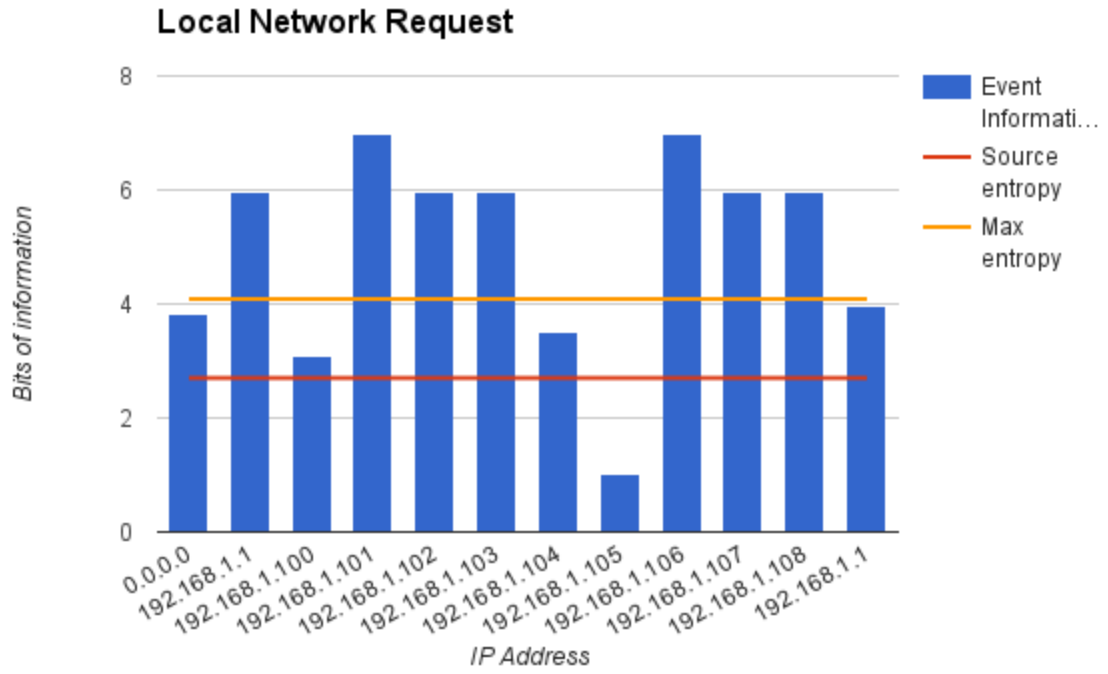
Local Network

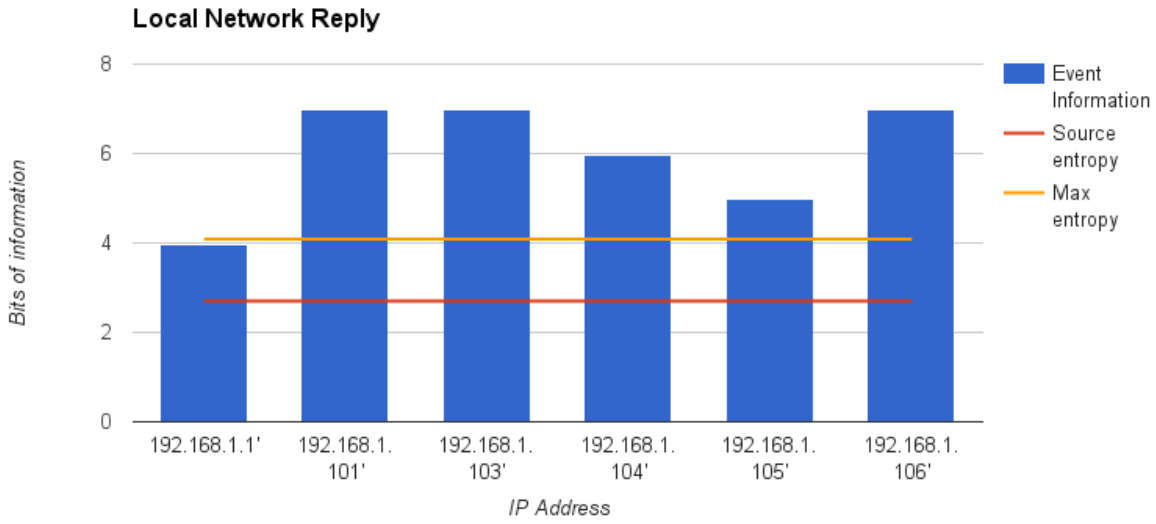




Se puede observar en estos gráficos que la red cuenta con claros nodos distinguidos. Para los paquetes de tipo *request*, vemos al nodo 192.169.1.105 como un nodo distinguido de la red, y la razón de esto es que ese disposivo particular era una computadora que estaba utilizando el servicio de *Whatsapp web,* que era hosteado por otro dispositivo móvil en la misma red.

El segundo gráfico nos distingue otro nodo en la red, que es quien efectuó la mayor cantidad de respuestas. Este nuevo nodo distinguido corresponde al *router*, siendo este el principal receptor de los mensajes *ARP Request*.

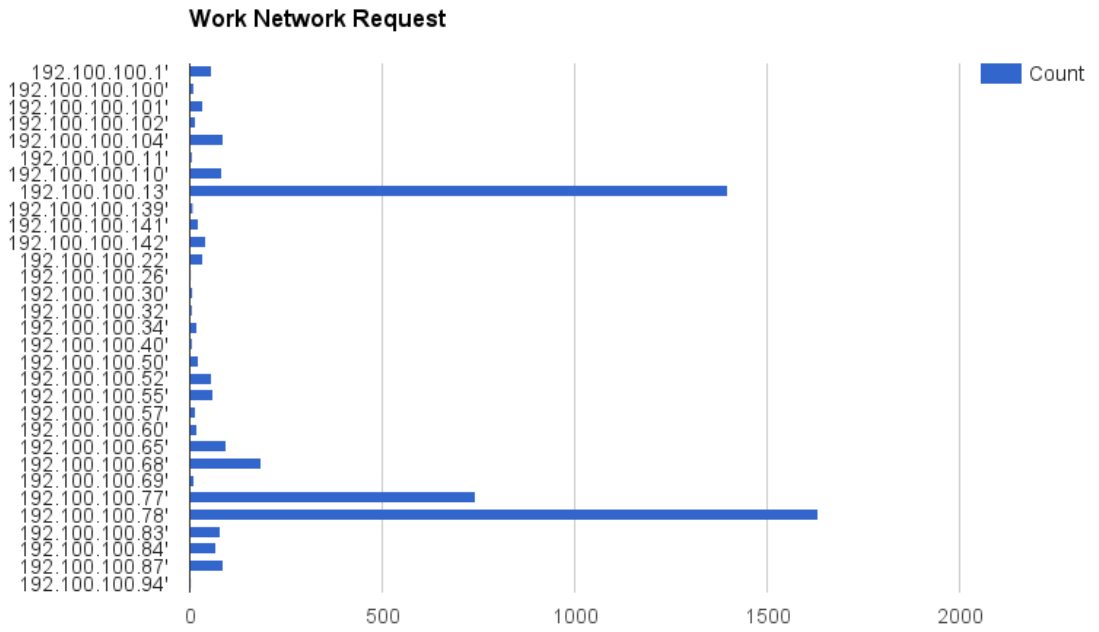


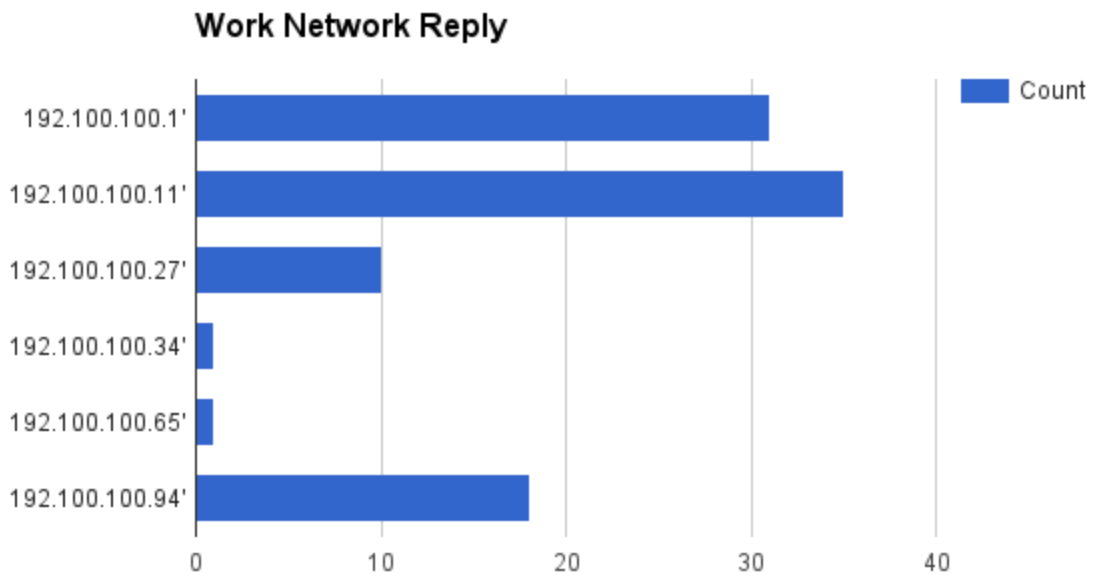


Podemos observar como efectivamente el nodo *192.168.1.105* que distinguimos se encuentran por debajo de la entropía de la red. Sin embargo, el *router* que identificamos como nodo distinguido cuando analizamos por separado los mensajes de tipo *request* y *reply*, analizandolo todo junto y comparándolo con la entropía de la red es significativamente mayor.

Work Network

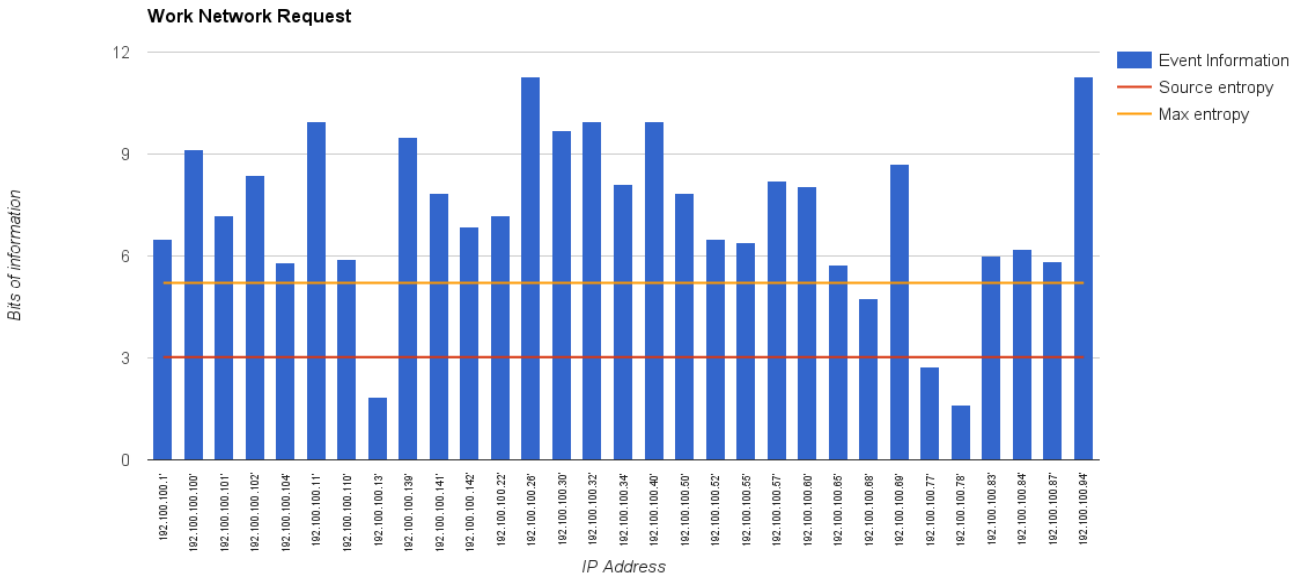
Para esta red, utilizamos gráficos de barra en vez de torta, ya que con el segundo, la información era poco clara, debido a la cantidad de nodos que contiene.

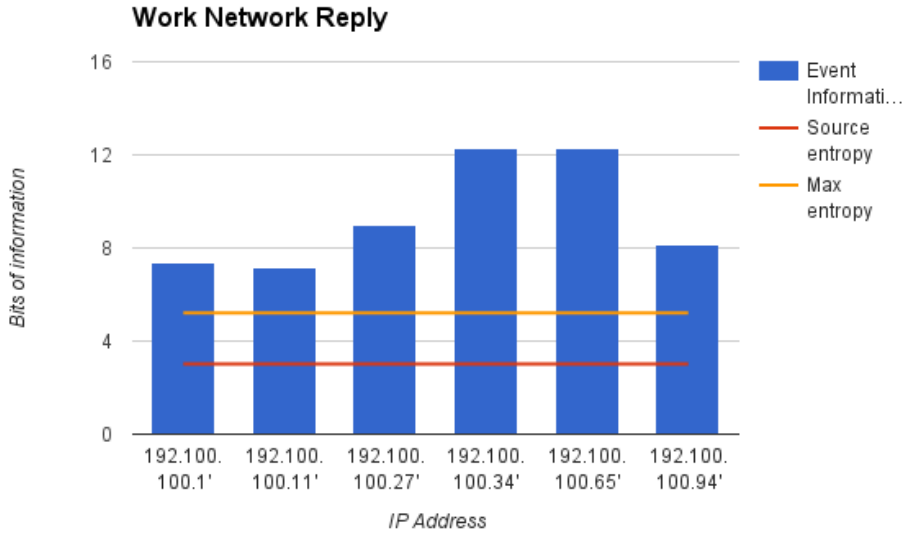




En el caso de los request, podemos notar como nodos distinguidos a las direcciones IP 192.100.100.13, 192.100.100.77 y 192.100.100.78. Dado el conocimiento que tenemos de la red, sabemos que estas direcciones corresponden con servidores de *testing*, las cuales tienen mucha utilización en el trabajo habitual de la oficina.

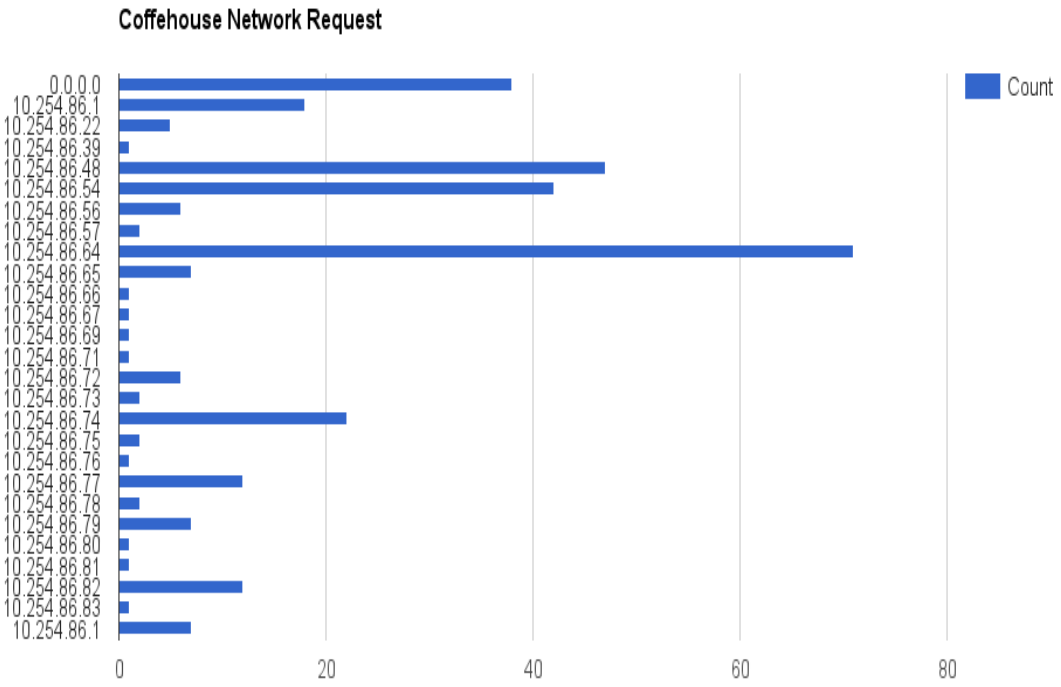
Por otro lado, en el *reply*, los nodos distinguidos corresponden primordialmente a 192.100.100.1 y 192.100.100.11, así como, en menor medida, 192.100.100.94. Sabemos que la primera dirección corresponde al *router*, la segunda a un servidor de base de datos y el tercero correspondiente al servidor del repositorio de fuentes, los cuales son claramente fundamentales para el trabajo diario.

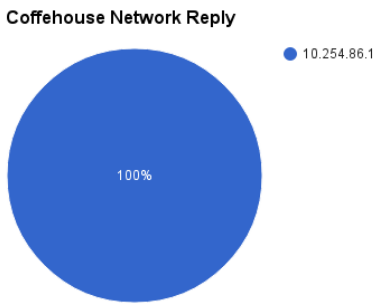




Se observa que, como se explicó anteriormente, los distintos servers recurrentes estan por debajo de la entropía de la red, aportando muy pocos bits de información en comparación a, por ejemplo, la dirección *192.100.100.50*, que es una estación de trabajo.

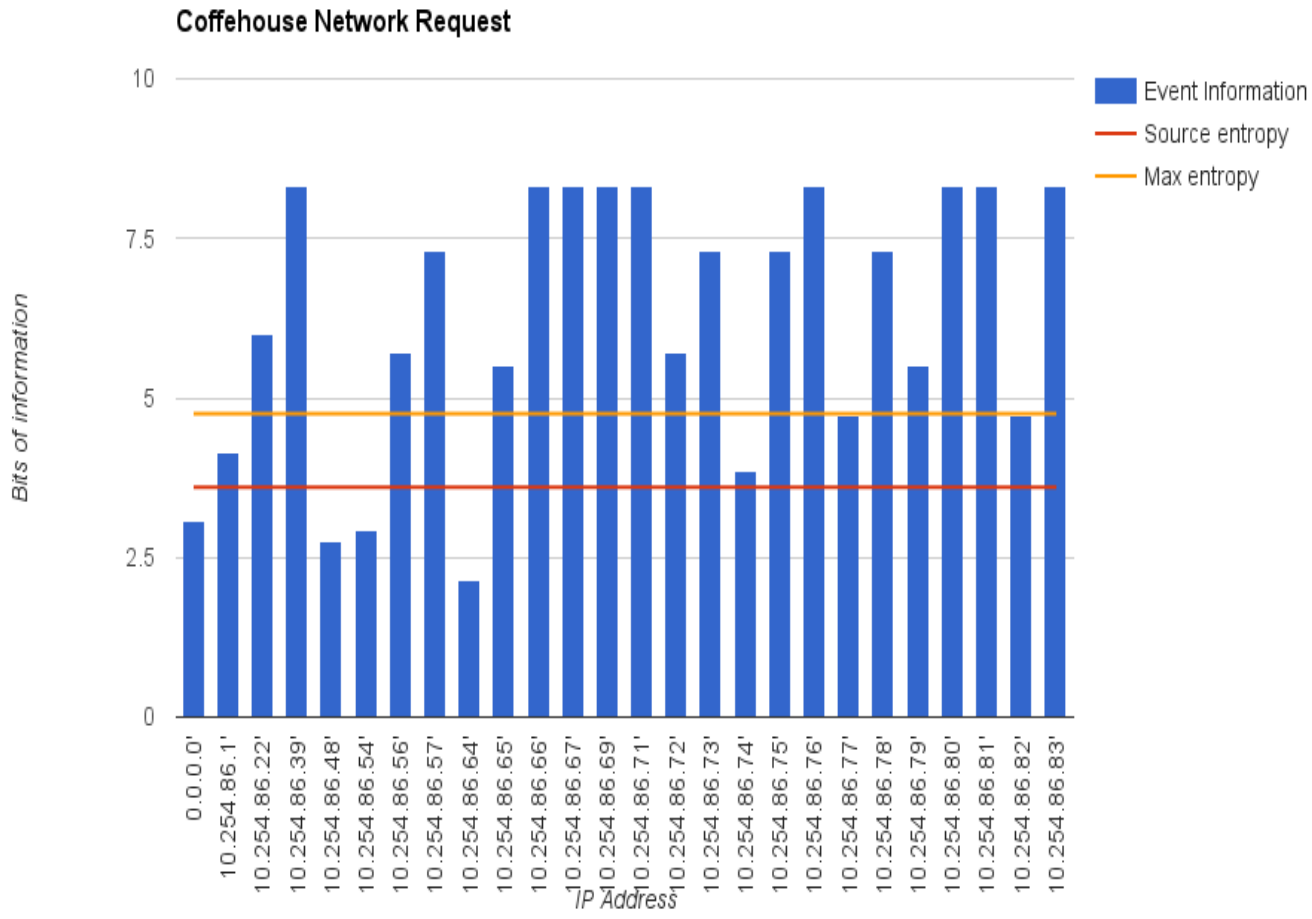
CoffeeHouse Network

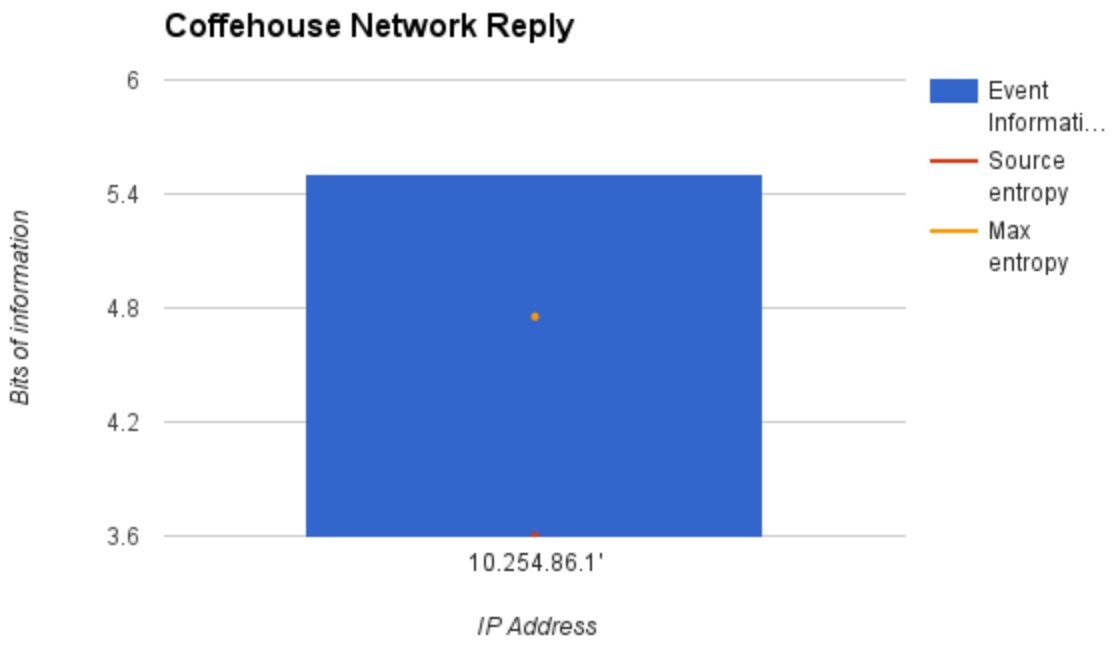




Por un lado, observamos que la cantidad de envíos de paquetes ARP *request* para la dirección IP 10.254.86.64 supera con creces a las demás. Esto lo convierte en un nodo distinguido y muy activo en la red. En menor medida, lo hacen las direcciones 10.254.86.48 y 10.254.86.54, resaltando también en la red. Sin embargo, estas direcciones muy posiblemente hayan sido dispositivos conectados a la red Wi-Fi del café, ya que todas las comunicaciones realizadas son contra el nodo central, el cual consideramos el *router*. El único caso donde una de estas tres IPs se conecta con alguien distinto del *router* es la dirección 10.254.86.54. Sin embargo, lo hace contra lo que llamamos *link local*, lo cuál implica un error en la conexión.

Ahora bien, consideramos *router* a la dirección IP 10.254.86.1 debido a ser el único nodo que envió ARP *reply*. Además, esta dirección fue *target* del código ARP para la mayor parte de los nodos de la red. Esto se puede visualizar claramente en la representación del grafo en el cual se lo muestra como nodo central a una esfera de nodos que se comunican a él. Podríamos pensar, a priori de más experimentación, que el router es un claro nodo distinguido.





Es interesante notar que el *router* (10.254.86.1) que parecía tan claramente un nodo distinguido en los grafos que se presentaron anteriormente, en cuanto se los compara con la entropía de la red se observa que está muy por encima de ésta, aún cuando resultó ser la única fuente de paquetes de tipo *ARP Reply*.

Shopping Mall Network

Los resultados obtenidos del tráfico de paquetes ARP de esta red no fueron considerados interesantes para este análisis. Sin embargo, sí cabe destacar el hecho de que en el corto tiempo de intervención la proporción de ARP fue muy baja y la red fue muy estable, contrario a lo que se hubiera pensado a priori.

Conclusiones Generales

En todas las redes, la entropía fue considerablemente menor que la máxima. La equiprobabilidad es un concepto teórico, que dista de las redes observadas. Más aún, si fuera efectivamente equiprobable la aparición de ARP e IP en las redes analizadas, se estaría enviando información redundante, disminuyendo el *troughput* de información relevante, y no de mantenimiento.

La red de trabajo, obtuvo el mayor porcentaje de ARP. Si fuera mucho más grande, y tuviera el mismo uso, el porcentaje de ARP sería aún mayor. Al agrandar la red, el costo de mantener la relación IP-MAC puede ser un limitante en su escalabilidad. Una posible solución, sería dividir en subredes, para limitar la inundación con tráfico de control.

En todos los casos, el *router* es un nodo distinguido de la red. Además, suele ser el principal emisor de *ARP-Reply*, dado que es el punto neurálgico de la comunicación entre los nodos de la red interna y el exterior.

Es necesario conocer la topología de la red, y su objetivo, para poder optimizar su funcionamiento. Hemos visto que redes con similar cantidad de dispositivos, poseen comportamientos completamente heterogéneos.

La red *CoffeeHouse* es el claro ejemplo de la existencia de un nodo distinguido, y crítico. Nuevamente, dependiendo del objetivo de la red, puede ser estrictamente necesaria la implementación de mecanismos de redundancia, para, ante la falla del nodo crítico, los usuarios puedan continuar con sus actividades.

Implementación

Para el desarrollo de esta investigación, se utilizó el lenguaje *Python,* con la biblioteca *Scapy*.

Generamos dos programas, el primero: para poder escuchar pasivamente los paquetes que se mueven en la red local. Permite también, guardar la información obtenida en un archivo con extensión *.cap/.pcap*. El fuente correspondiente se encuentra bajo el nombre *passive_listener.py*.

El segundo, para procesar las fuentes de información de la cátedra y la ideada por el grupo, como también generar las estadísticas. Su fuente se encuentra en *source_analizer.py*. Las estadisticas generar un archivo *.csv*. Los gráficos se hicieron con herramientas de hoja de cálculo que procesan este tipo de archivo.

Modo de uso

- **Passive listener:** *python passive_listener.py [listening time in minutes][name of dump]*. El primer parámetro es el tiempo en el cual se va a estar escuchando, el segundo el nombre del archivo donde se guardará la información recopilada.

- **Source analizer:** *python source_analizer.py [analysis source: type|arp|graph][source file][statistics file]*. El primer parámetro define que tipo de análisis se realizará sobre la fuente:
    - Type: Obtendrá la información sobre la fuente definida por la cátedra.
    - ARP: Obtendrá la información sobre la fuente definida por nosotros.
    - Graph: Mostrará las relaciones entre los nodos.

  El segundo y tercer parámetro, definen el nombre de archivo fuente, y dónde se guardarán las estadísticas respectivamente.

Limitaciones

Scapy no funciona correctamente en computadoras con *Windows*, o por lo menos, no lo pudimos lograr. Si bien el código de ambos programas funcionan sin arrojar error, en el caso de la escucha pasiva, muchos paquetes recibidos, después no pueden ser analizados. Abriéndolos tanto en *WireShark,* como en nuestra implementación, estos paquetes aparecen como *RAW*.

No pudimos generar en estas computadoras una máquina virtual con *Linux*: no tienen activada la virtualización por hardware, y el acceso a la *bios* está restringido.

Decidimos entonces validar que nuestra implementación funcione igual que *WireShark* para una entrada pequeña, y luego usamos ambos programas indistintamente para generar los *datasets* de información. La muestra del *Shopping* fue creada con nuestro

programa, y las restantes con *WireShark*. La creación de las fuentes y estadística, fue realizada enteramente con *source_analizer.py*.

Vale la pena aclarar, que por la arquitectura en la que el programa de análisis corría, no pudimos analizar fuentes de más de 170.000 símbolos. Dado que el *Scapy* instalado corría con *Python* de 32 bits, una instancia de *Python* no puede ocupar más de 2GB. Procesando más de 185.000 símbolos, el programa se quedaba sin memoria, y era desalojado por el sistema operativo. Dado que las fuentes en algunos casos, eran más grandes, decidimos acotar el proceso de *sniff*, dejando 100MB para el procesamiento restante.

Entregable

El entregable consta del informe en formato pdf, una carpeta *src* donde podrán encontrar el código fuente, una carpeta *muestras_procesadas* donde encontrarán los archivos con toda la información del análisis y una carpeta llamada *grafos* donde podrán encontrar los archivos *.graphml* que contienen los grafos. Estos archivos se abren con el programa YED[7]. Utilizándolo, podrán ver el detalle y navegar la topología de la red.

[7] http://www.yworks.com/en/products/yfiles/yed/