Teoría de las Comunicaciones
TP1

20 de abril de 2016

Integrante	LU	Correo electrónico
Martin Baigorria	575/14	martinbaigorria@gmail.com
Federico Beuter	827/13	federicobeuter@gmail.com
Mauro Cherubini	835/13	cheru.mf@gmail.com

Reservado para la cátedra

Instancia	Docente	Nota
Primera entrega		
Segunda entrega		

Índice

1. Introducción

La comunicación es uno de los ejes fundamentales de la humanidad. A lo largo de la historia han aparecido diferentes medios para poder satisfacer esta necesidad, en 1948, el matemático e ingeniero Claude E. Shannon propone una definición formal de que es la comunicación desde una punto de vista matemático, donde origen a la *Teoría de la Información*. En este trabajo analizaremos como aplica dicha teoria a un medio de comunicacion real, particularmente uno que sea altamente utilizado y tenga una alta densidad de usuarios.

2. Desarrollo

2.1. Informacion y Entropia

Como vimos en la introduccion, en 1948 el matematico Claude E. Shannon define formalmente que es la *Informacion* en su publicacion *A Mathematical Theory of Communication*, junto con esta defincion tambien introduce el concepto de *Entropia* en la comunicacion. Primero definimos quienes son los participes en la comunicacion:

- Fuente de Informacion
- Emisor
- Receptor
- Destino de Informacion
- Ruido
- Canal

Los mismos se encuentran conectados de la siguiente forma:

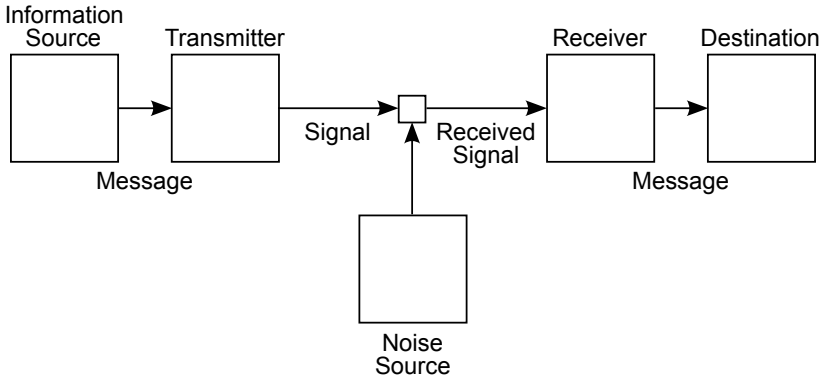

Figura 1: Diagrama de comunicacion

Para poder establecer una comunicacion punto a punto, es necesario que el emisor y el receptor «hablen» un lenguaje en comun. Para poder satisfacer esto, se define un conjunto de simbolos $S = \{s_1, ..., s_n\}$, con una probabilidad $p(s_i)$ con $1 \leq i \leq n$ asociada a cada uno de ellos, este conjunto representa la totalidad de simbolos que pueden ser transmitidos por el canal. Una vez definido el conjunto, la informacion que se obtiene por simbolo es simplemente $I(s_i) = log(\frac{1}{p(s_i)})$. Se elige el logaritmo como funcion por cumplir con varias condiciones idóneas para calcular la informacion, de las mas interesantes tenemos que:

- Si $s \in S$, entonces $I(s) \geq 0$
- Si $p(s_i) = 1$ para algun i, entonces $I(s_i) = 0$, ya que un evento que ocurre siempre no aporta información significativa
- $I(s_i, s_j) \leq I(s_i) + I(s_j)$, la igualdad vale unicamente si los simbolos son independientes

Como podemos apreciar, la funcion logaritmo cumple con todos estos puntos.

En nuestro trabajo analizaremos principalmente la entropia, esta se define como $H(S) = \sum_{i=1}^{n} p(s_i)I(s_i)$. Esta medida representa la media de informacion obtenida por simbolo en la comunicacion, y se encuentra íntimamente relacionada

con las probablidades de cada simbolo, particularmente mientras menos aleatoria sea una red, menor es la entropia. Tambien aplica el mismo razonamiento a la inversa, es decir, mientras mas cerca de la equiprobabilidad esten los simbolos, la entropia aumentara, maximizándose cuando los simbolos sean equiprobables.

2.2. Canal elegido

Para poder tener suficientes datos para hacer un analisis interesante, seria prudente tomar un medio que sea ampliamente usado. Por ello, elegimos utilizar una red del tipo Wi-Fi, los puntos a destacar de la red son:

- En la aplicaciones habituales, estas redes se caracterizan por tener un *nodo* central, el cual se encarga de regular el trafico en la red
- En las redes publicas, los nodos que no son el central no se suelen comunicar entre ellos, estos principalmente se conectan con servidores en internet

2.3. Paquetes de red

En la redes la informacion en el canal se transmite en paquetes, estos paquetes tienen varios campos, particularmente nos interesan analizar los paquetes de capa 2 y capa 3. De los paquetes *Ethernet* de capa 2, nos interesa:

- Direccion MAC de origen
- Direccion MAC de destino
- Protocolo del payload, este depende del paquete de capa 3 que se esta transportando, puede ser IPv4, IPv6, ARP, etc.

Por otro lado, de los paquetes de capa 3 nos interesan solamente los ARP. De este tipo de paquetes nos interesan los campos:

- Tipo de operacion
- Direccion MAC del emisor
- Direccion IP del emisor
- Direccion MAC del destinatario
- Direccion IP del destinatario

El fin de los paquetes ARP, es vincular la capa 2 con la capa 3, relacionando las direcciones IP con direcciones MAC fisicas. Para hacer esto el protocolo ARP cuenta con dos operaciones, estas pueden ser *who-has* o *is-at*.

Las operaciones *who-has* sirven para identificar a que direccion MAC fisica corresponda una direccion IP de la red, estos paquetes suelen ser de tipo *broadcast*.

En respuesta a los *who-has*, tenemos las operaciones *is-at*. Una vez que un nodo recibe un paquete de tipo *who-has*, este revisa si la direccion IP del mismo coincide con la suya, en el caso de que lo sea este envia un paquete al nodo que genero el *who-has* para notificarle su direccion MAC fisica.

Para no tener que enviar paquetes ARP por cada paquete IP que se desea enviar, el emisor del *who-has* al recibir el *is-at*, guarda el resultado en una tabla para poder enviar futuros paquetes al mismo destino inmediatamente.

3. Metodologia

En el trabajo se pide que analicemos dos fuentes de informacion, estas son S y S_1, y tienen la siguiente forma:

- S: Comprende a todos los paquetes *Ethernet* que circulan por el canal, se los diferencia por el *protocolo* del payload.
- S_1: Se limita a los paquetes *Ethernet* de tipo ARP.

Para S_1 ademas se pide establecer un criterio de diferenciacion, para esto tomamos la conjuncion de las direcciones IP fuente y destino de los paquetes ARP de tipo *who-has*. Elegimos estos paquetes ya que los *is-at* se dan unicamente en respuesta a algun *who-has* anterior, con lo cual estariamos duplicando ciertos paquetes.

Las redes elegidas para *sniffear* fueron las siguientes:

- FibertelZone, en shopping Plaza Oeste Moron (60 min. de captura)
- Laboratorios-DC (30 min. de captura)

Inicialmente la idea era capturar paquetes en una mayor cantidad de redes, sin embargo, nos topamos con que en diferentes redes publicas el trafico no era suficientemente alto, con lo cual la captura de paquetes no era significativa ni suficiente como para hacer un analisis mas exhaustivo. Consideramos que esto se debe en parte al aumento en el uso de Smartphones con redes moviles, junto con la mala fama que tienen las redes Wi-Fi publicas.

4. Resultados

4.1. Protocolos

Vamos a ver primero la cantidad de paquetes en las redes:

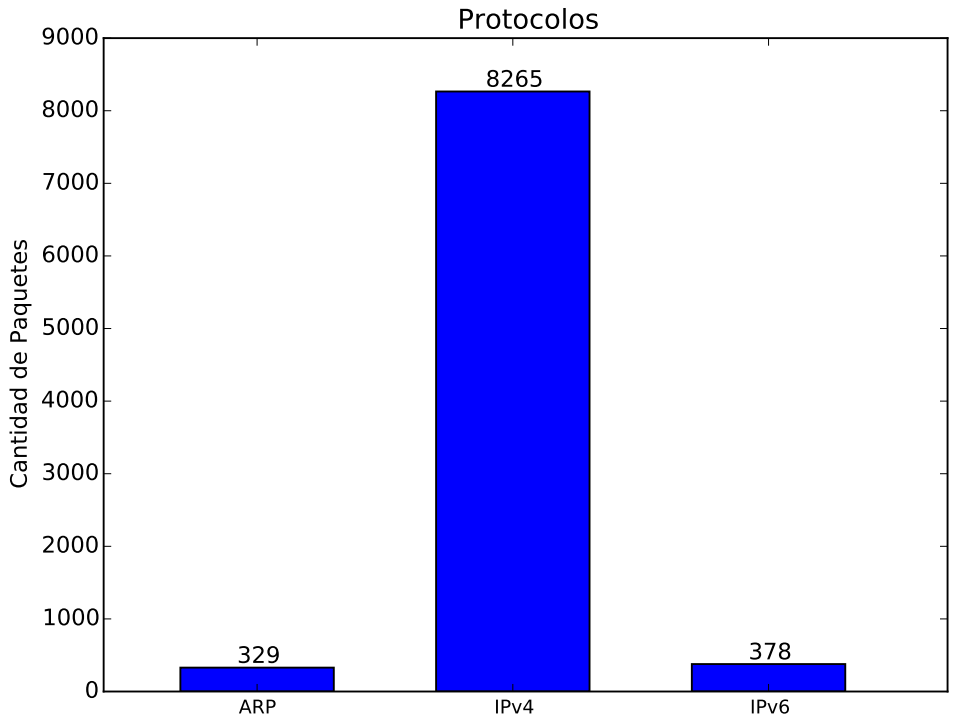

Figura 2: Red Plaza Oeste

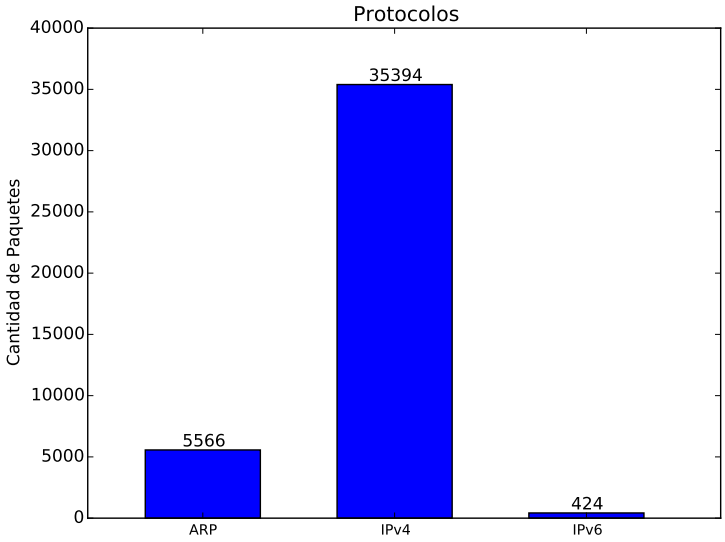

Figura 3: Red Laboratorios DC

Como podemos apreciar, los paquetes IPv4 dominan el trafico de la red. A continuacion vamos a ver la informacion por simbolo junto con la entropia:

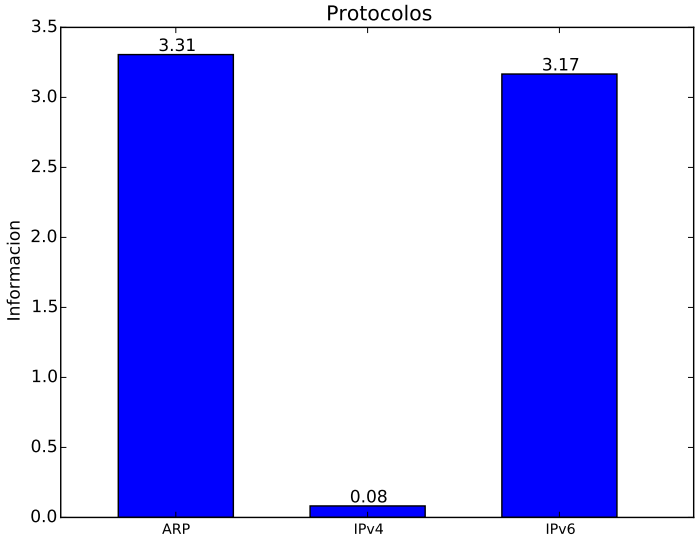

Figura 4: Red Plaza Oeste

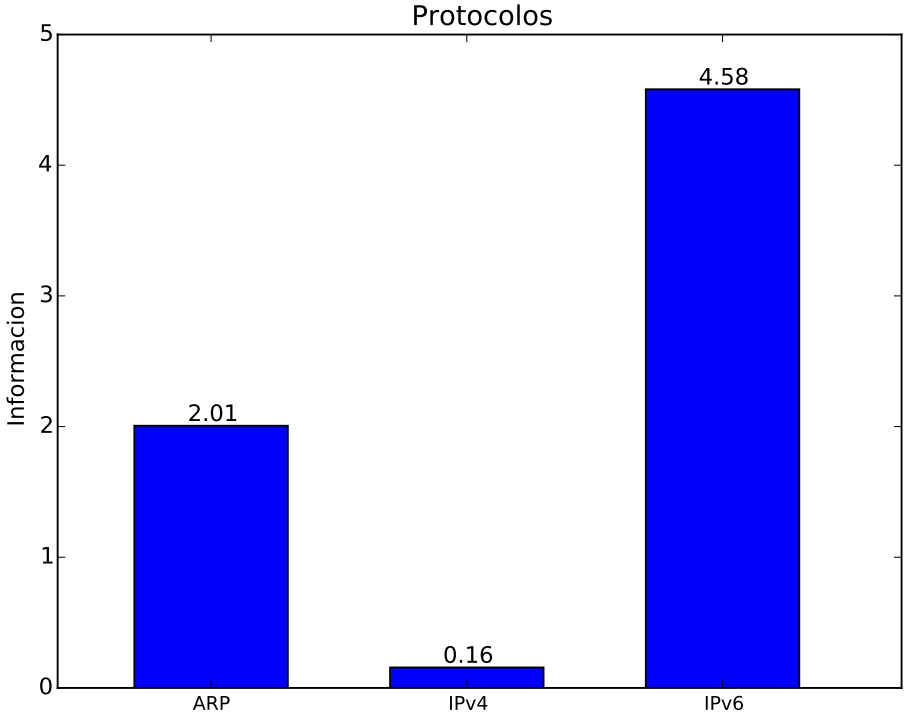

Figura 5: Red Laboratorios DC

La entropia en las redes fue:

Cuadro 1: Entropia

Red	Entropia
Plaza Oeste	0.3303
Laboratorios DC	0.4505

Como podemos ver, la entropia de la red es bastante baja en ambos, esto se debe a la gran cantidad de paquetes IPv4 en ambos canales. Otro punto interesante es que en la red unicamente circulan paquetes de tipo ARP, IPv4 o IPv6, si bien esto es esperable en la red del Plaza Oeste nos sorprendió que ocurra también en la red Laboratorios DC, asumimos que allí podríamos encontrar en uso algún protocolo mas esotérico.

Nuestras predicciones indicaban que la mayoría de los paquetes iban a ser de tipo IPv4 por un margen bastante significativo, esto se corresponde con que la probabilidad de dicho paquete iba a ser mucho mas alta que el resto, con lo cual la información de dichos paquetes iba a ser considerablemente menor. Como la diferencia es tan amplia, por un margen muy significativo, consideramos al protocolo IPv4 como símbolo distinguido.

4.2. Paquetes ARP

Como vimos en el desarrollo, definimos los símbolos de S_1 como las direcciones IP de origen y destino de los paquetes ARP. Por motivos de presentación, nos limitamos a graficar unicamente las diez direcciones que mas aparecieron durante el *sniffeo*. Los resultados fueron:

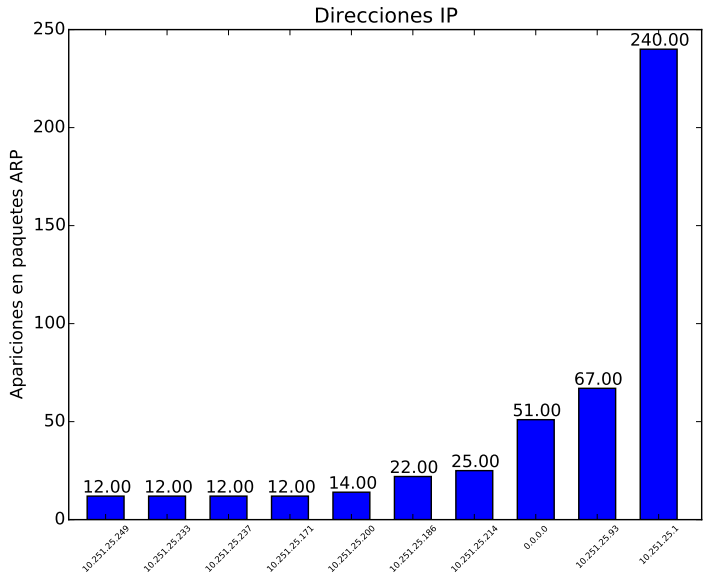

Figura 6: Red Plaza Oeste

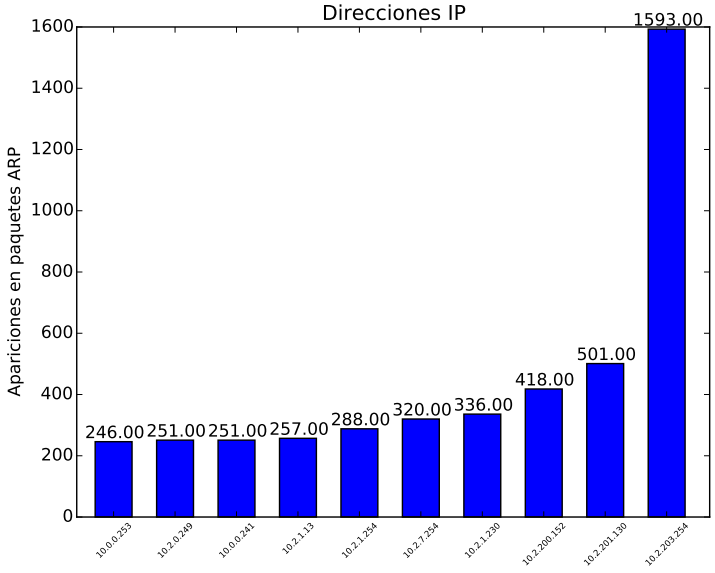

Figura 7: Red Laboratorios DC

Una de las suposiciones que teníamos, era que por el funcionamiento de la red Wi-Fi, la gran mayoría de los paquetes irían dirigidos hacia un nodo principal, el cual después se encargaría de redirigirlos al destino apropiado. Como podemos ver, esto ocurrió en ambos casos, hay un nodo el cual tiene muchas mas apariciones que el resto.

A pesar de que los resultados satisfacían nuestras teorías, nos sorprendió que el margen de diferencia entre el nodo principal y el resto no sea mayor, con lo cual los otros nodos también están enviando paquetes ARP a nodos que no son el principal. Esto sera estudiado mas adelante, primero vamos a ver la información en ambos canales, nuevamente limitándonos a las diez direcciones que mas aparecieron:

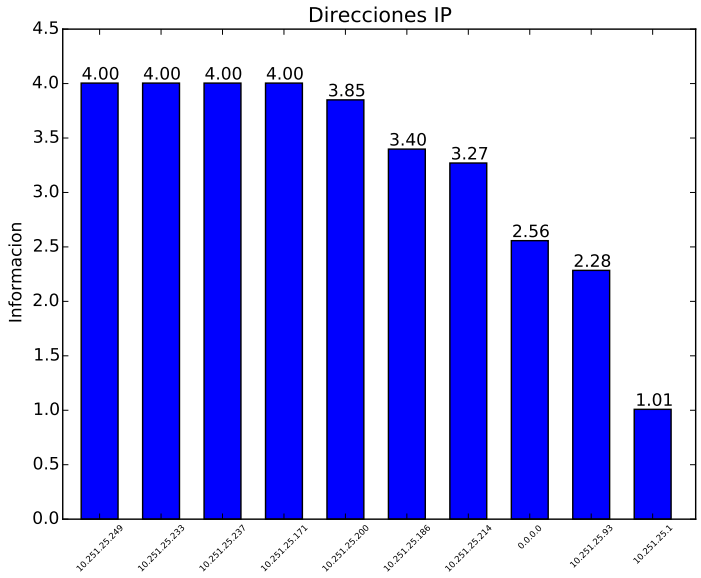

Figura 8: Red Plaza Oeste

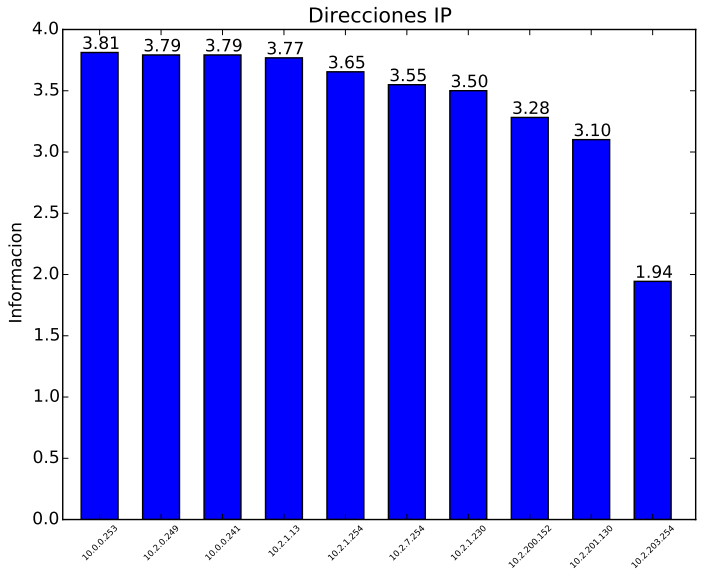

Figura 9: Red Laboratorios DC

La entropia para estas dos redes fue:

Cuadro 2: Entropia

Red	Entropia
Plaza Oeste	3.1724
Laboratorios DC	4.6459

Una cosa interesante a destacar, es que la entropia es mucho mayor tomando como símbolos las direcciones IP que los tipos de protocolos, esto ocurre en ambas redes. Esto creemos que se debe principalmente a las siguientes razones:

- La cantidad de protocolos es considerablemente menor que la cantidad de direcciones IP posibles, en la practica unicamente vimos tres protocolos en uso
- En las redes Wi-Fi, vimos que los nodos que suelen enviar paquetes de tipo *who-has* son aquellos que quieren conectarse a la red, estos suelen tener direcciones IP diferentes a las que ya se encuentran en el sistema, con lo que tenemos que agregar un nuevo símbolo
- Es posible que el tamaño de la muestra no haya sido suficiente, y que se necesiten mas tiempo de *sniffeo* para poder calcular bien la entropia

Este ultimo punto es particularmente interesante para nosotros, ya que es posible que la muestra tomada no sea significativa del trafico de la red y no sea suficiente para estimar las probabilidades, efectivamente afectando la entropia de la red. También seria interesar estudiar si las políticas de asignación de IP de la red y el uso de la misma terminan amortizando el segundo punto, ya que si el sistema reasigna direcciones siempre que puede, mientras mas grande sea la muestra la diferencia entre el nodo principal y el resto potencialmente seria mayor.

En general, la red respondió de la manera que esperábamos, pudimos identificar un nodo principal el cual aparece en muchas mas ocasiones que el resto en ambas redes. Si bien la diferencia en cantidad de apariciones de dichos nodos respecto al resto no fue tan significativa, creemos que fue suficiente como para marcarlos como símbolos distinguidos en sus respectivos canales.

4.3. Paquetes de control ARP

En la sección anterior hablamos de paquetes que no eran enviados hacia el nodo principal, ademas de esto, nos dimos cuenta que varios de ellos tenian una forma bastante particular, tras consultar diferentes recursos nos dimos cuenta que varios de ellos eran paquetes de control. Nos topamos con los siguientes:

- Dirección IP `0.0.0.0`: Estos paquetes se utilizan para revisar si una dirección IP se encuentra en uso por algún host, la idea es que un nodo al recibir una dirección IP para usar, envia este paquete, y si recibe un *is-at* de otro nodo quiere decir que la dirección que pretendía utilizar esta en uso
- Misma dirección IP de fuente y destino: Este es un paquete bastante particular, sirve para que los diferentes hosts en la red tengan sus tablas de dirección IP y MAC actualizadas. La idea es que al recibir el paquete y verificar que las direcciones origen y destino son iguales, el host revisa si tiene la dirección MAC en su tabla, en caso de tenerla actualiza la dirección IP almacenada si es que esta cambio por la dirección que figura en el paquete. Mantener las relaciones IP y MAC actualizadas es sumemente importante en la red, ya que eso puede ahorrarnos una cantidad significativa de tiempo a la hora de manejar el trafico.

Las apariciones de estos paquetes consideramos que terminaron quitandole bastante peso al nodo central respecto a los demás, haciendo que la entropia sea mayor.

5. Conclusiones

Es bastante interesante ver como los diferentes conceptos de teoria de la informacion aplican en canales reales, particularmente considerando que fueron planteados en 1948 cuando las redes modernas no existian. Desde el punto de vista de las redes Wi-Fi puntualmente, vimos como la comunicacion es sumamente centralizada y dominada por IPv4, lo primero se puede ver claramente en los paquetes ARP los cuales en nuestras muestras estan en gran parte dirigidos a un unico nodo, mientras que lo segundo no solamente se ve a simple vista, sino que ademas es esperable considerando que la Internet funciona sobre IPv4.

Como estudio a futuro, seria interesante hacer un *sniffeo* durante el lapso de uno o mas dias, para poder ver si las proporciones obtenidas con la fuente S_1 se mantienen, o si eventualmente el nodo principal se impone respecto al resto aumentando aun mas la diferencia entre estos.