

Trabajo Práctico Nº1: Wiretapping

Alvaro Jose Fernando, Barbeito Nicolás, Brum Raúl, Nieves Yésica

Resumen—Our premise is ...

I. INTRODUCCIÓN

A. Paquetes ARP

El protocolo ARP (Address Resolution Protocol) permite mapear direcciones de nivel de red a direcciones físicas. La idea de este protocolo se basa en el envío de paquetes que pueden ser de preguntas o respuestas. El emisor del paquete que pregunta por una dirección, envía un mensaje broadcast sobre la red local, siendo respondido por un mensaje unicast por aquel al que pertenece la dirección consultada. Mediante el envío de estos paquetes ARP se construyen las tablas que mapean direcciones de red con direcciones físicas.

B. Entropía de una fuente

Para poder definir la entropía de una fuente, necesitamos la definición de información que aportan los símbolos emitidos por dicha fuente. Se define información de un símbolo s como

$$I(s) = \log(1/P(s))$$

siendo $P(s)$ la probabilidad de ocurrencia de dicho símbolo. De esta manera, puede calcularse la información media suministrada por una fuente de información de memoria nula (los símbolos emitidos son estadísticamente independientes) como

$$\sum_S P(s_i) I(s_i) \forall s_i \in S$$

Esta cantidad media de información por símbolo de la fuente, recibe el nombre de *entropía* $H(S)$ de la fuente de memoria nula.

$$H(S) = \sum_S P(s_i) \log(1/P(s_i))$$

Debido a que la entropía de una fuente depende de la probabilidad de los diferentes símbolos que la componen, se puede demostrar que para una fuente de información de memoria nula con un alfabeto de q símbolos, el valor máximo de la entropía es precisamente $\log q$, alcanzándose solamente si todos los símbolos son equiprobables.

C. Fuente S

Sea P la fuente de información generada a partir de todos los paquetes Ethernet que se transmiten en una determinada red entre los instantes de tiempo $[t_i, t_f]$:

$$P_{ti,tf} = \{p_1 \dots p_n\}$$

siendo p_i el i -ésimo paquete transmitido en la red entre los instantes de tiempo $[t_i, t_f]$.

Alvaro Jose Fernando, LU: 89/10, email: fer1578@gmail.com
Barbeito Nicolás, LU: 147/10, email: nicolasbarbeiton@gmail.com
Brum Raúl, LU: 199/98, email: brumraul@gmail.com
Nieves Yésica, LU: 340/05, email: yesica.nieves@gmail.com

Los paquetes p_i pertenecientes a P encapsulan diferentes protocolos, que se pueden identificar a través del campo type del frame de capa 2 (p.type en Scapy). Por lo tanto, con el objetivo de distinguir los protocolos utilizados en una red, se define otra fuente de información S de la siguiente manera:

$$S_{ti,tf} = \{s_1 \dots s_n\}$$

siendo $s_i = p_i.type / p_i$ perteneciente a P entre los instantes de tiempo $[t_i, t_f]$.

D. Propuesta de una nueva fuente S_1

Para analizar la entropía de la red en base a los paquetes ARP observados realizamos una nueva tool en base a la anterior que nos permitiera obtener datos de los campos de dichos paquetes. Se propone como nueva fuente S_1 el conjunto de símbolos conformado por las distintas direcciones IP destino:

$S_1 = \{s_{11} \dots s_{1n}\}$ siendo s_{1i} el valor del campo *pdst* correspondiente a la ip destino del paquete

Al igual que para la fuente S , realizamos el cálculo de la entropía como fue requerido, como la probabilidad e información de sus símbolos.

II. MÉTODOS Y CONDICIONES DE CADA EXPERIMENTO

En esta sección describiremos brevemente las redes elegidas para realizar las escuchas mediante las herramientas indicadas en el enunciado del trabajo practico.

A. Home Lan

Esta medición fue realizada en la Lan de una casa por un intervalo de 2 horas. La misma cuenta con una computadora corriendo un sistema operativo Linux la cual es el router de la Lan y provee a las demás computadoras de acceso a internet ademas de otros servicios de red (proxy, dns, dhcpd, etc). A la misma se encuentran conectadas mediante un switch 4 computadoras cableadas y 2 access point inalámbricos. A estos últimos se encontraban conectados al momento de la medición una notebook y varios celulares. Ademas una de las computadoras cableadas corre una maquina virtual con ip propia independiente y otra de las computadoras cableadas posee otro isp para conectarse a internet por lo que no utiliza a la primera computadora como gateway.

B. Labos Dc

Esta medición fue realizada en los laboratorios del departamento de computación por un intervalo de 30 minutos. Se desconoce la cantidad de computadoras y celulares conectados al momento de realizar la captura.

C. Zona pública (parque)

Se realizó una medición en un parque con wifi abierto durante 15 minutos. Debido a que la medición fue realizada en un espacio público se espera que varios dispositivos móviles como celulares y notebooks se conecten a la misma, sin poder precisar cuántos a priori.

D. Universidad (Universidad de la Matanza)

Esta medición se realizó por aproximadamente 10 minutos en la red wifi de la Universidad de La Matanza. Es una red abierta y desconocemos la estructura de ella.

III. RESULTADOS Y ANÁLISIS

En esta sección mostraremos y analizaremos los resultados obtenidos en las mediciones realizadas en las distintas redes tanto para la fuente S y S_1 descritas en la sección I C y D. Para cada una de las redes veremos sus protocolos distinguidos, la proporción de paquetes ARP en el tráfico de la red y los nodos (representados por ip) distinguidos. Cabe aclarar que para la fuente S_1 podemos ver todos los pedidos “who has” de ARP ya que todas las maquinas están en el mismo dominio de broadcast pero solo podremos ver las respuestas “is at” de aquellas maquinas que se encuentren en igual dominio de colisión (generalmente las conectadas por wireless ya que las cableadas al estar conectadas por switches no tienen colisión y solo el destinatario recibirá la respuesta).

A. Home Lan

A.1 Fuente S

Los resultados obtenidos para la fuente S fueron:

Protocolo	Información	Probabilidad
EAPOL	12.31	0.01%
ARP	5.11	2.88%
IPv6	4.22	5.33%
IPv4	0.12	91.76%

TABLE I
HOME LAN - PROTOCOLOS

Como puede observarse en la figura 1 el protocolo mas utilizado es IPv4 en un 91.76% mientras que IPv6 y ARP solo son utilizados en un 5.33% y 2.88% respectivamente. EAPOL (autenticación wireless) no tiene prácticamente incidencia. Observamos ademas que el protocolo ARP tiene solo una incidencia del 2.88% en el trafico total de la fuente lo que hace que el overhead aportado por el mismo no sea significativo.

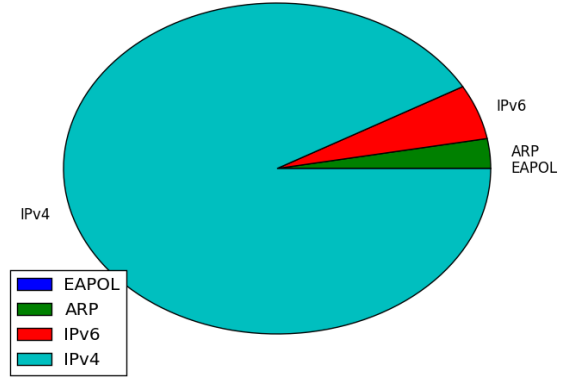


Figura 1. Home Lan - Probabilidades

Por otro lado la entropía de la fuente S fue de 0.4892 lo que hace que los símbolos emitidos por la fuente S sean muy previsibles. Esto podemos notarlo en la figura 2 donde observamos que el protocolo con mayor porcentaje de aparición, IPv4, sea el que menos información aporta a la fuente. La información aportada por este se encuentra por debajo de la entropía de S . Como contraste observamos en la figura 2 que el protocolo EAPOL es el que mas información aporta pero según lo observado en la figura 1 tiene una probabilidad muy baja lo que hace que no incida en la entropía.

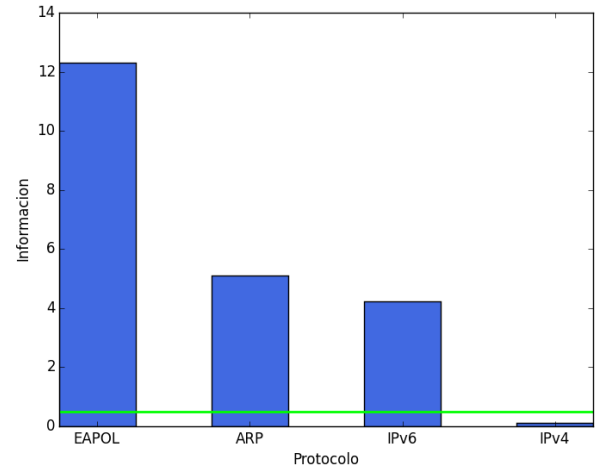


Figura 2. Home Lan - Información

A.2 Fuente S_1

Para la fuente S_1 expondremos las ip de destino, los resultados fueron :

Protocolo	Información	Probabilidad
192.168.10.10	7.68	0.48%
192.168.10.27	6.41	1.16%
192.168.10.11	6.30	1.26%
192.168.10.13	5.09	2.92%
192.168.10.7	4.33	4.97%
192.168.10.1	4.00	6.23%
192.168.10.4	2.54	17.15%
192.168.10.6	2.11	23.09%
192.168.10.9	1.22	42.69%

TABLE II
HOME LAN - NODOS

Como puede observarse en la figura 3 las ip que aparecen con mayor frecuencia como destino en los paquetes ARP son la “192.168.10.9” en un 42.69%, la “192.168.10.6” en un 23.09% y la “192.168.10.4” en un 17.15% por lo que estas serán los nodos distinguidos de la fuente y las que menos información aporten.

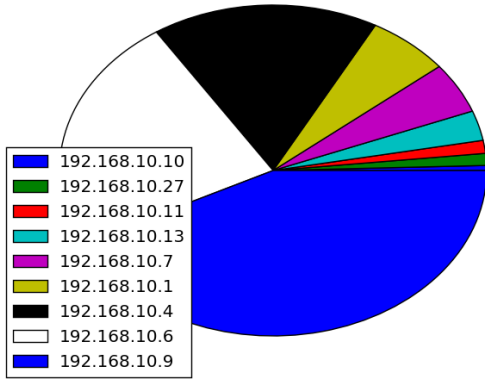


Figura 3. Home Lan - Probabilidades

La entropía de la fuente S_1 fue de 2.25. En la figura 4 se observa la información aportada por los distintos nodos y como 2 de los nodos distinguidos quedan por debajo de la entropía de la fuente y el otro apenas por arriba.

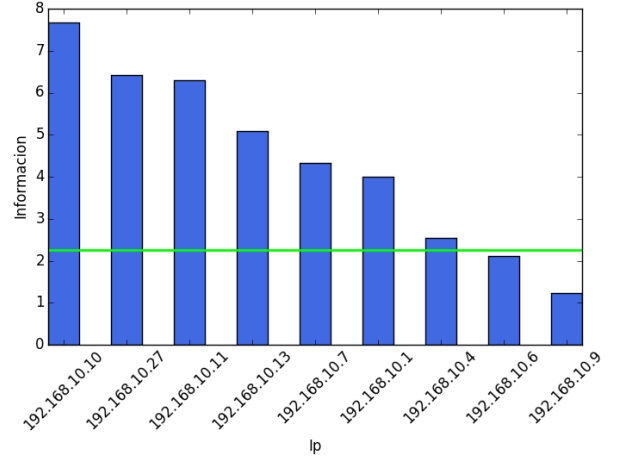


Figura 4. Home Lan - Información

En el análisis de la fuente S_1 vemos que ninguno de los 3 nodos distinguidos pertenecen al router de la red, la ip “192.168.10.1”. De los nodos distinguidos las ip “192.168.10.6” y “192.168.10.4” se explican ya que la primera es la notebook que realizo la captura en la red y la segunda es la computadora que la controlaba remotamente mediante SSH. La ip “192.168.10.9” corresponde a la ip de una consola de video juegos la cual se encontraba apagada al momento de realizar la captura, en la ip “192.168.10.4” hay un daemon ejecutándose que se comunica con la consola de video juegos. Al estar esta apagada hay muchos pedidos “who has 192.168.10.9” y ninguna respuesta por lo que los pedidos son reiterados varias veces incrementado su incidencia.

B. Labos Dc

B.1 Fuente S

Los resultados obtenidos para la fuente S fueron:

Protocolo	Información	Probabilidad
ARP	8.092	0.37%
IPv4	0.005	99.63%

TABLE III
LABOS DC - PROTOCOLOS

Como observamos en la figura 5 ampliamente el protocolo mas utilizado es IPv4 en un 99.63% mientras que ARP solo es utilizado en un 0.37%. El overhead aportado por que el protocolo ARP en el trafico de la red despreciable. Notamos también la ausencia de protocolo para wireless aunque habian celulares conectados a la red durante la captura. Creemos que esto es debido a que la red wifi provista por los labos del Departamento de Computación es abierta y al no estar encriptadas no aparece el protocolo EAPOL utilizado durante la autenticación.

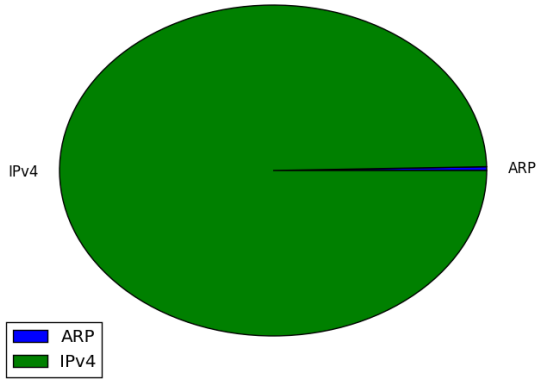


Figura 5. Labos Dc - Probabilidades

La entropía de la fuente S fue de 0.034922 lo que hace que los símbolos emitidos por la fuente S sean muy previsibles como podemos notarlo en la figura 6 donde observamos que la información aportada por el protocolo IPv4 se encuentra por debajo de la entropía de S . El protocolo ARP aporta información muy por arriba de la entropía como se observa en la figura 6 pero según lo observado en la figura 5 tiene una probabilidad demasiado baja lo que hace que no incida en la entropía.

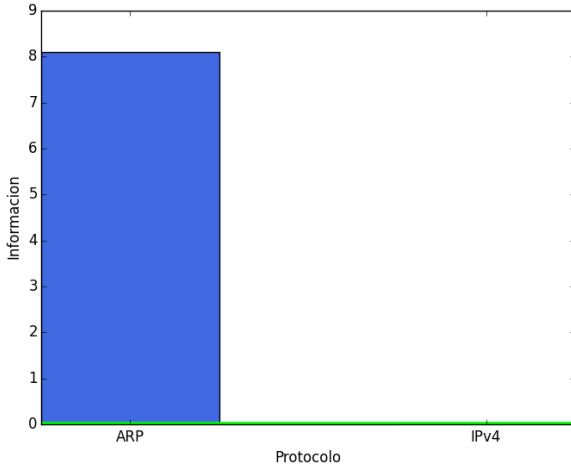


Figura 6. Labos Dc - Información

B.2 Fuente S_1

Para la fuente S_1 expondremos las ip de destino. Dala la cantidad de ips capturadas para facilitar la lectura de los gráficos y tablas expondremos 2 bloques (A yB) en representación de ips de destino que tuvieron igual aporte de información en la captura. Para el caso del bloque A se compone de 19 ip con igual aporte de información y el bloque B de 7 ips. Los resultados fueron :

Protocolo	Información	Probabilidad
A(19 ips)	8.40	5.60%
B(7 ips)	7.40	4.13%
10.2.203.144	6.82	0.88%
10.2.202.5	6.82	0.88%
10.2.202.228	6.82	0.88%
10.2.203.122	6.40	1.17%
10.2.203.117	6.40	1.17%
10.2.203.254	4.82	3.53%
10.2.201.214	4.23	5.30%
10.2.202.132	4.08	5.89%
10.2.200.217	4.01	6.19%
10.2.202.108	2.33	19.76%
10.2.202.100	2.33	19.76%
10.2.202.47	2.01	24.77%

TABLE IV
LABOS DC - NODOS

Como puede observarse en la figura 7 las ip que aparecen con mayor frecuencia como destino en los paquetes ARP son la “10.2.202.47” en un 24.77%, la “10.2.202.100” en un 19.76% y la “10.2.202.108” en un 19.76% por lo que estas serán los nodos distinguidos de la fuente y las que menos información aporten.

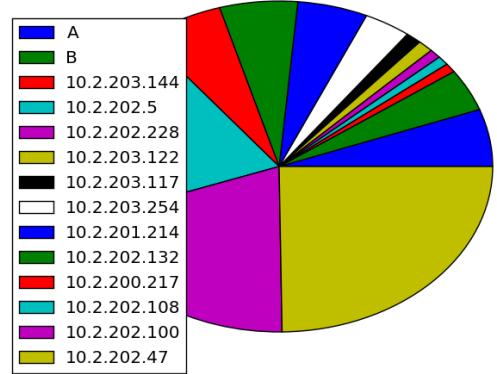


Figura 7. Labos Dc - Probabilidades

La entropía de la fuente S_1 fue de 3.417440. En la figura 8 se observa la información aportada por los distintos nodos y como los 3 nodos distinguidos quedan por debajo de la entropía de la fuente.

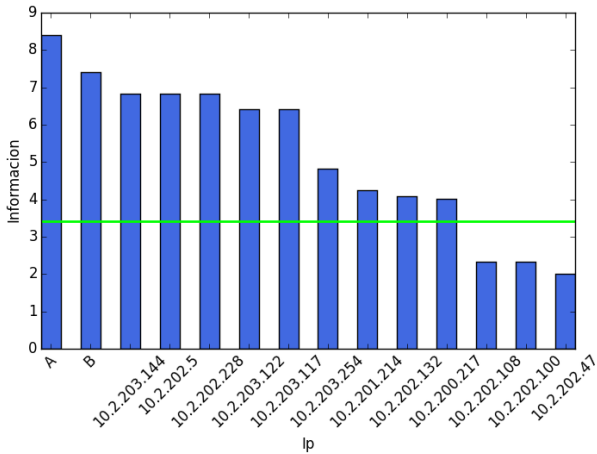


Figura 8. Labos Dc - Información

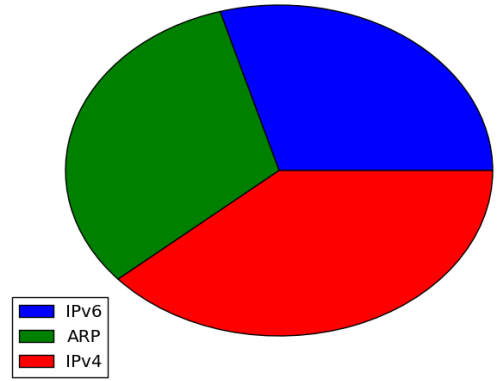


Figura 9. Zona Pública (parque) - Probabilidades

En el análisis de la fuente S_1 vemos que ninguno de los 3 nodos distinguidos pertenecen al router de la red, la ip “10.2.203.254” ni tampoco a la ip de la maquina que realizo la captura “10.2.202.132”. No tenemos información de la red para precisar a que pertenecen las ip de los nodos distinguidos.

En cuanto a la entropía de la fuente S , fue de 1,575466, lo que nos dice que los símbolos emitidos por la fuente no son previsibles, ya que ningún protocolo otorga mucha más información que los demás (ver figura 10). No podemos decir que para esta fuente haya un símbolo distinguido en esta fuente.

C. Zona pública (parque)

C.1 Fuente S

Los resultados obtenidos para la fuente S fueron:

Protocolo	Información	Probabilidad
IPv6	1.763	29.45 %
ARP	1.646	31.95 %
IPv6	1.373	38.59 %

TABLE V
ZONA PÚBLICA (PARQUE) - PROTOCOLOS

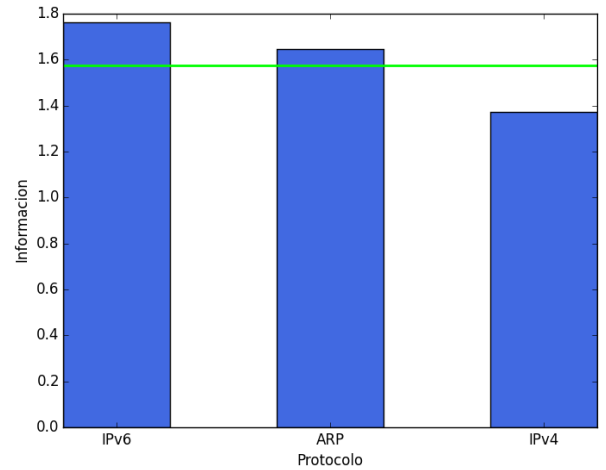


Figura 10. Zona Pública (parque) - Información

Podemos notar en la figura 9 que el protocolo más utilizado es IPv4 (38.5919%) seguido por ARP (31.9503%) e IPv6 (29.4578%). En comparación con las redes analizadas en secciones anteriores, los paquetes ARP son muy frecuentes. Otra peculiaridad es que el protocolo IPv4 y el IPv6 son usados en casi la misma frecuencia, lo que no sucedió en las redes anteriores.

C.2 Fuente S_1

Para la fuente S_1 expondremos las ip de destino. Dada la cantidad de ips capturadas, para facilitar la lectura de los gráficos y tablas, expondremos las direcciones ip que tuvieron igual aporte de información en la captura en bloques.

Protocolo	Información	Probabilidad
B1(171 ips)	14.105417	0.97%
B2(227 ips)	13.105417	2.56%
B3(348 ips)	12.422595	6.38%
B4(253 ips)	11.601712	8.25%
B5(40 ips)	10.649047	2.53%
B6(16 ips)	9.504669	2.25%
B7(4 ips)	8.569523	1.06%
B8(5 ips)	7.543969	2.69%
B9(4 ips)	6.495808	4.49%
169.254.255.255	5.380904	2.39%
10.249.147.254	4.970991	3.18%
10.248.79.254	4.908201	3.33%
10.248.0.100	4.340546	4.93%
10.248.0.68	4.277281	5.15%
10.249.111.254	4.129570	5.71%
10.248.255.254	3.972275	6.37%
169.254.188.151	3.521395	8.70%
10.249.107.249	3.278075	10.30%
10.250.127.254	2.422862	18.64%

TABLE VI
ZONA PUBLICA (PARQUE) - NODOS

La entropía de la fuente fue de 5,757719, lo que indica que la fuente tiene un alto nivel de imprevisibilidad. Podemos ver que los bloques de direcciones de B1 a B14 son nodos diferenciados debido a la alta cantidad de información aportada, es decir, su baja probabilidad de aparecer como símbolo en la fuente; de la misma manera, las direcciones ip 10.249.107.249 y 10.250.127.254 también se diferencian pero por la razón contraria, aportan muy poca información, por tener una frecuencia relativa de aparición alta durante las capturas en relación a los otros símbolos.

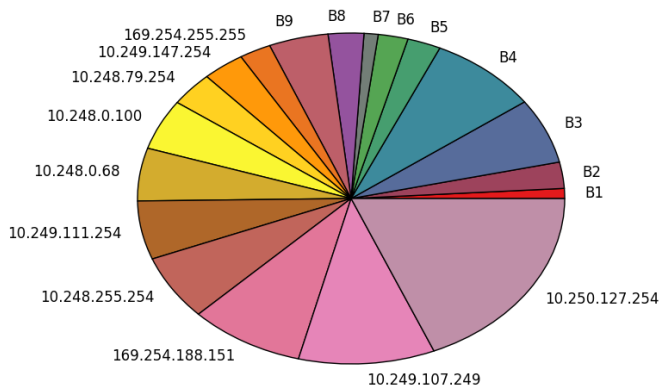


Figura 11. Parque S1 - Probabilidades

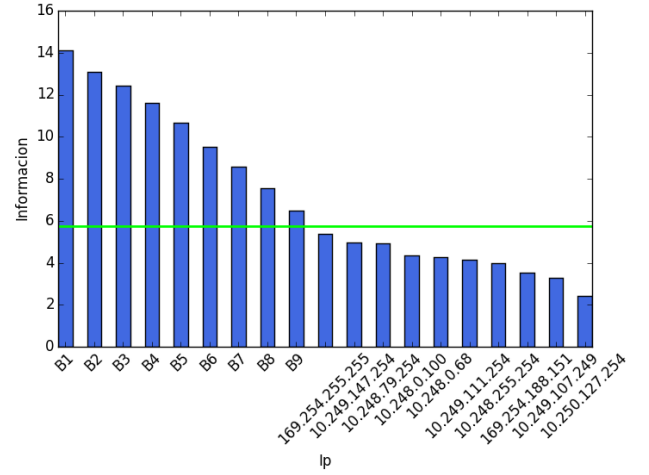


Figura 12. Parque - Información

Como particularidad en esta red observamos la presencia de nodos con ips fuera del rango usado por la red como ser la ip 169.255.188.151. Estas ip pertenecen a nodos que no tienen una ip asignada en la red y estan usando como ip la asignada por "ipv4ll" mediante "zeroconf". Esta tecnologia asigna una ip automáticamente cuando el nodo no puede obtener una ip a traves de DHCP o no tiene configurada una ip fija. Las ips asignadas de esta manera pertenecen al bloque de ips 169.254.0.0/16.

D. Universidad de La Matanza

D.1 Fuente S

Los resultados obtenidos para la fuente *S* fueron:

Protocolo	Información	Probabilidad
IPv6	6.753	0.009%
ARP	1.269	0.415%
IPv4	0.796	0.576%

TABLE VII
UNIVERSIDAD DE LA MATANZA - PROTOCOLOS

Se observa en la figura 13 que los protocolos más utilizados son IPv4 en un 57.60% y ARP con un 41.50%, el uso del protocolo IPv6 es de apenas 0.9

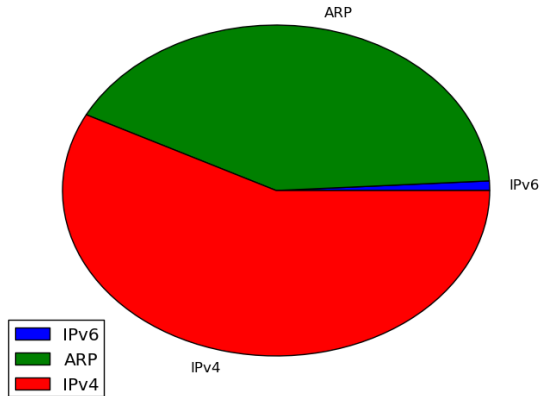


Figura 13. Universidad de La Matanza - Probabilidades

La entropía de la fuente S fue de 1.047737 lo que hace que los símbolos emitidos por la fuente S sean muy imprevisibles como podemos notar en la figura 14 donde observamos que la información aportada por el protocolo IPv4 se parece a la del protocolo ARP ambos muy cerca de la entropía de S . El protocolo IPv6 aporta información muy por arriba de la entropía según la figura 14, pero según lo observado en la figura 13 tiene una probabilidad demasiado baja lo que hace que no incida en la entropía.

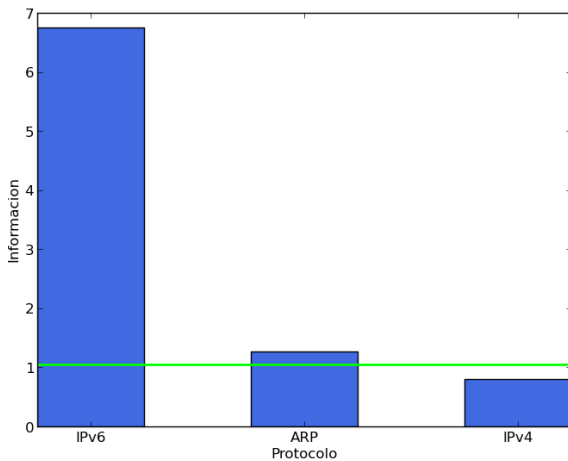


Figura 14. Universidad de La Matanza - Información

D.2 Fuente S_1

Para la fuente S_1 expondremos las ip de destino. Dada la cantidad de ips capturadas para facilitar la lectura de los gráficos y tablas expondremos varios bloques de (A hasta D) en representación de ips de destino que tuvieron un aporte de información parecido en la captura. Como había demasiadas ips, para las que estaban muy por encima de la entropía se utilizó un promedio para darle un valor para el gráfico. Los resultados fueron :

Protocolo	Información	Probabilidad
A(428 ips)	11.27	0.0005%
B(92 ips)	9.48	0.0014%
C(56 ips)	8.48	0.0029%
10.5.81.167	8.00	0.0039%
10.5.84.65	7.93	0.0041%
D(25 ips)	7.65	0.0050%
10.5.83.154	7.00	0.0079%
10.5.78.34	6.94	0.0081%
10.5.77.136	6.87	0.0086%
10.5.84.63	6.58	0.0105%
10.5.70.216	6.12	0.0144%
10.5.200.13	5.91	0.0166%
10.5.0.1	5.53	0.0216%

TABLE VIII

UNIVERSIDAD DE LA MATANZA - NODOS

Como puede observarse en la figura 15 las ip que aparecen con mayor frecuencia como destino en los paquetes ARP son la “10.5.0.1” en un 2.16%, la “10.5.200.13” en un 1.66%, la “10.5.70.216” en un 1.44% y la “10.5.84.63” en un 1.05% por lo que estas serán los nodos distinguidos de la fuente y las que menos información aporten.

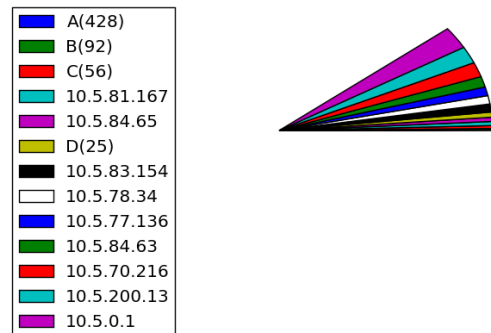


Figura 15. Universidad de La Matanza - Probabilidades

La entropía de la fuente S_1 fue de 7.951183. En la figura 16 y la figura 15 se observa la información aportada por los distintos nodos y como los 4 nodos distinguidos quedan por debajo de la entropía de la fuente. En este caso había muchas ips por debajo de la entropía (casi 30 ips), así que tomamos como punto de corte los que tengan una probabilidad mayor a 1

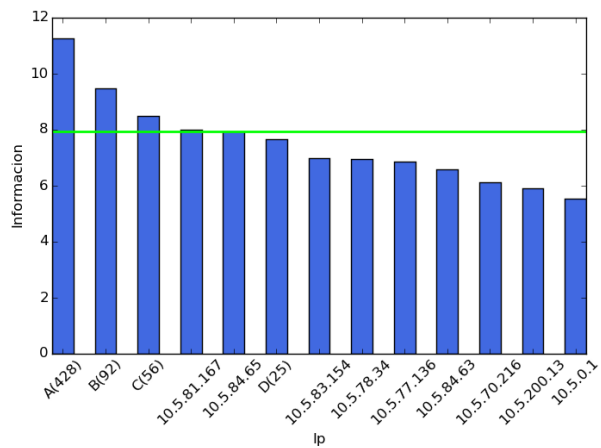


Figura 16. Universidad de La Matanza - Información

En el análisis de la fuente S_1 vemos que la ip “10.5.0.1” podría corresponderse algún router y de los otros 3 nodos no disponemos de suficiente información sobre la red para saber cual es el rol que cumplen.

IV. CONCLUSIÓN

En este trabajo se analizaron 4 redes distintas desde el punto de vista de la teoría de la información. Este enfoque nos permitió distinguir y cuantificar la importancia de distintos entes dentro de las redes, ya sean protocolos y su uso relativo en las redes LAN analizadas o nodos que las conforman. (FALTA UNA RED PARA LAS CONCLUSIONES)

V. REFERENCIAS