DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico Nª 1

Teoria de las comunicaciones

Segundo Cuatrimestre de 2017

Integrante	LU	Correo electrónico
Juan Ignacio Noli Villar	174/14	juaninolivillar@gmail.com
Axel Lew	225/14	axel.lew@hotmail.com
Matias Cadaval	345/14	matias.cadaval@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. Introducción

El objetivo del actual trabajo es utilizar técnicas provistas por la teoría de la información para estudiar diversos aspectos de una red de manera analítica.

Nos enfocaremos en este trabajo practico en el análisis de las capas dos y tres de la red.

La capa dos, la capa de enlace de datos se ocupa del direccionamiento físico, del acceso al medio, de la detección de errores, de la distribución ordenada de tramas y del control del flujo.

La capa tres, la capa de red tiene como objetivo hacer que los datos lleguen desde el origen al destino, aun cuando ambos no estén conectados directamente.

Analizaremos tres redes con diferentes características, y a partir del análisis de los paquetes que se envían a través de la red, llegaremos a conclusiones sobre el estado de las mismas. Serán redes de diferente tamaño, con diferente trafico, conectándonos por Ethernet o por Wi-Fi. Para cada red obtendremos una cantidad grande de paquetes, para que nuestro análisis sea lo menos propenso a errores y outliers posible.

Utilizamos para esto dos herramientas. Wireshark, utilizada para obtener los paquetes que son transmitidos en la red dada y Scapy, para el procesamiento y análisis de los paquetes.

Un primer análisis constara en ver cuales son los protocolos de la capa superior inmediata de los paquetes, y que tipo de destino tienen (es decir, broadcast o unicast).

Para esto utilizaremos las herramientas antes mencionadas, y a partir de ellas, obtendremos información sobre la red, como la entropía muestral de los protocolos.

En segundo lugar, desarrollamos una herramienta, hecha en Python, con el fin de identificar los hosts de nuestra red.

Se define la entropía como la incertidumbre de una fuente de información. Más formalmente,

$H(S) = -\sum_i p(s_i) * log_2(p(s_i))$

Utilizaremos la entropía en nuestro análisis, dado que nos permite distinguir símbolos sobre otros

2. Capturado y modelado del tráfico

2.1. Ejercicio 1

Se programo una herramienta usando python y la librería Scppy para poder analizar .pcaps. Para utilizar la misma simplemente ejecutar en la terminal:

```
$ python3 main.py <ruta del pcap>
```

Definimos como fuente de información nula $S_1 = \{s_1, ..., s_n\}$ con s_i formado por la combinación entre el tipo de destino de la trama y el protocolo de la capa inmediata superior encapsulado en la misma.

La herramienta procesara cada paquete y luego mostrara por consola

- Entropía de la fuente
- Entropía máxima
- Ocurrencia y probabilidad de cada símbolo de la fuente

2.2. Ejercicio 2

Para poder identificar los host de una captura primero se filtran todos los paquetes que no utilicen el protocolo ARP.

El protocolo ARP es un protocolo de comunicaciones de la capa de enlace, responsable de encontrar la dirección de hardware (Ethernet MAC) que corresponde a una determinada dirección IP.

Los paquetes resultantes se modelaran de la siguiente manera:

```
Simbolo = <paquete.SRC>
```

Luego utilizaremos el siguiente criterio:

- Los paquetes ARP con mas ocurrencia son los que pertenecen al router, pues el mismo debe saber en todo momento que MAC posee cada host conectado a la red
- Los paquetes que le siguen en ocurrencia pertenecen a los host de la red
- Los paquetes con menos ocurrencia vienen de Internet, o cualquier otra red externa

Entonces, bajo este criterio, los paquetes con mas ocurrencia son los hosts (excepto el primero, que probablemente sea el router). Por definición de información de un evento, estos son los que menos información tienen.

Entonces nos quedamos con los símbolos que tengan menos información que la información promedio de la fuente. De esos paquetes, asumimos que la IP que se encuentre en SRC pertenece a un host

2.3. Ejercicio 3

Utilizando el modelo planteado en el ejercicio 2, adaptamos la herramienta para que lo utilice y así identifique a los hosts de una red. Ahora, además de mostrar todo lo pedido en el ejercicio 1, también muestra por consola los hosts identificados.

3. Consigna 2: Gráficos y análisis

A continuación se explicaran los experimentos llevados a cabo. Para los mismos se utilizó la herramienta desarrollada para la consigna 1.

Para identificar los hosts se utilizo el modelo planteado para la resolución del ejercicio 2.

Graficaremos para nuestro análisis la red como un grafo dirigido. Habrá un nodo por cada IP participe de la red y una arista de a y b si existe un paquete ARP con a como fuente y b como destino.

Dado a la gran cantidad de nodos (es decir, IPs que han mandado o recibido un paquete ARP) decidimos unir ciertos nodos con el fin de simplificar el grafo. Para esto haremos los siguiente. Para cada conjunto maximal de nodos tal que ninguno haya mandado un ARP y todos reciban únicamente de un mismo nodo, los agruparemos en un solo nodo, cuyo valor sera la cantidad que representa.

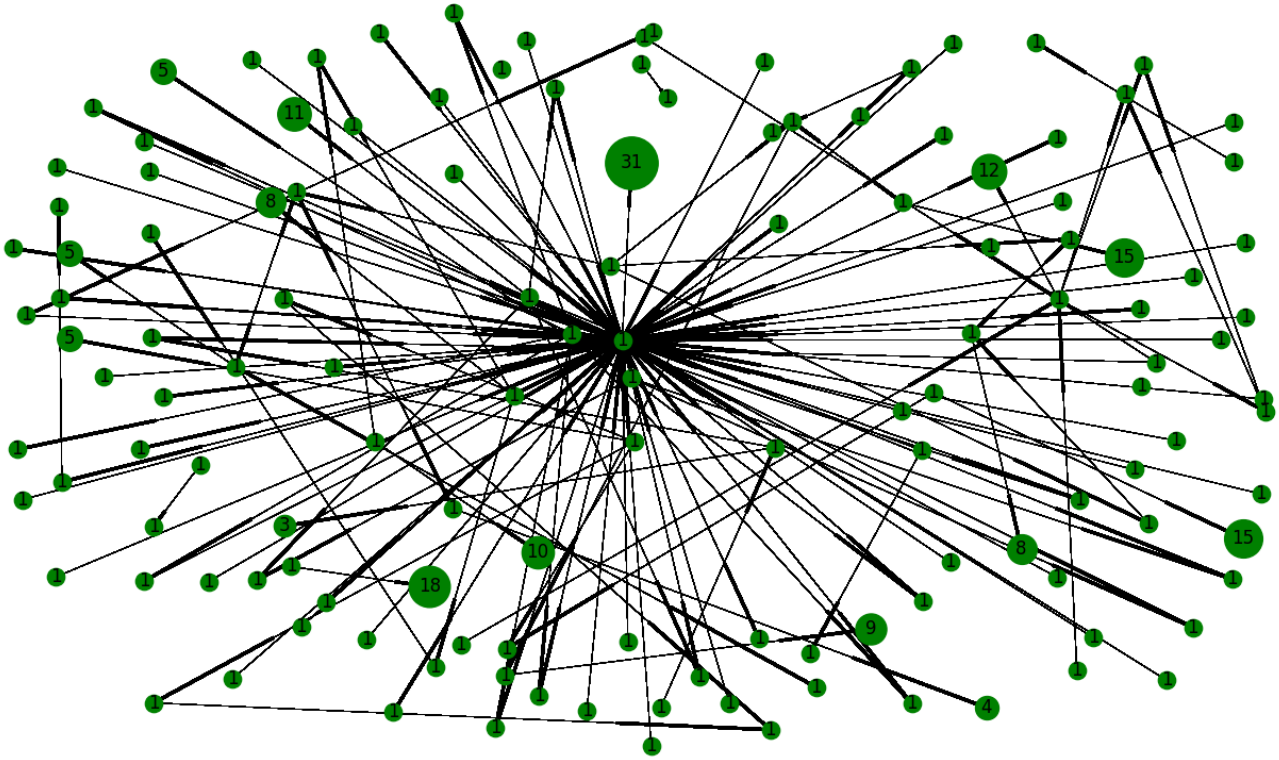
4. Laboratorios Dc

4.1. Condiciones

Para el siguiente experimento se utilizara una captura realizada en los laboratorios del DC.

- Tamaño: Medio
- Horario: 5pm
- Duración: 15 minutos
- Modo: Ethernet

4.2. Grafo de la red



4.3. Porcentaje de protocolos

```
Probabilidades de protocolos:

'IP':       57,5,
'ARP':      13,4,
'LLC':      5,4,
'EAPOL':    0,3,
'IPv6':     22,3,
'Raw':      1,0

Porcentaje de paquetes broadcast: 36.99
```

4.4. Información de los paquetes

```
Cantidad total de paquetes: 10992
Entropía:                   2.2504489157535
Entropía Máxima:            3.321928094887362

symb: ('UNICAST',   'IP')       count: 3876     prob: 0.35262
symb: ('BROADCAST', 'ARP')      count: 1478     prob: 0.13446
symb: ('UNICAST',   'LLC')      count: 470      prob: 0.04275
symb: ('UNICAST',   'EAPOL')    count: 29       prob: 0.00263
symb: ('UNICAST',   'IPv6')     count: 2455     prob: 0.22334
symb: ('BROADCAST', 'IP')       count: 2445     prob: 0.22243
symb: ('BROADCAST', 'Raw')      count: 14       prob: 0.00127
symb: ('BROADCAST', 'LLC')      count: 129      prob: 0.01173
symb: ('UNICAST',   'Raw')      count: 94       prob: 0.00855
symb: ('UNICAST',   'ARP')      count: 2        prob: 0.00018
```

4.5. Hosts distinguidos

```
average_info:       8.979708951242175

IP                  Info
'10.2.0.195'        5.0719498418790145
'10.2.201.120'      7.0719498418790145
'10.2.203.254'      1.8589561185448165
'10.2.0.198'        5.20945336562895
'10.2.0.64'         5.105116705814214
'10.2.0.65'         5.105116705814214
'10.2.0.67'         5.105116705814214
'10.2.0.254'        4.67340046538874
'10.2.0.191'        5.20945336562895
'10.2.201.184'      6.0719498418790145
'10.2.4.6'          7.0719498418790145
'10.2.6.250'        6.624490864907793
'10.2.2.250'        6.624490864907793
'10.2.203.7'        6.83094174237522
'10.2.203.100'      6.946418959795156
'10.2.203.68'       6.624490864907793
'10.2.203.104'      6.946418959795156
```

4.6. Análisis

La herramienta detecto varios símbolos distinguidos, 17 para ser mas precisos. Esto tiene sentido, pues la red ethernet de un laboratorio del dc posee al rededor de 20 equipos.

Observando la probabilidad de los protocolos, se puede deducir que uno de los principales usos de la red es transmitir datos, ya que la probabilidad de que un símbolo sea ('UNICAST', 'IPv6') o ('UNICAST', 'IP') es de casi un 50 %

Por otro lado, la probabilidad de ('BROADCAST', 'IP') no es menor, por lo que se puede decir que se suelen transmitir los mismos datos a todas las computadoras (se puede asociar al hecho de que la información entre todas las computados de los labos se encuentra sincronizada).

Se puede notar que la entropía no esta tan alejada de la entropía máxima. Esto se debe a que , como observamos anteriormente, en esta red se utilizan ambas tramas de destino (BROADCAST Y UNICAST) casi por igual.

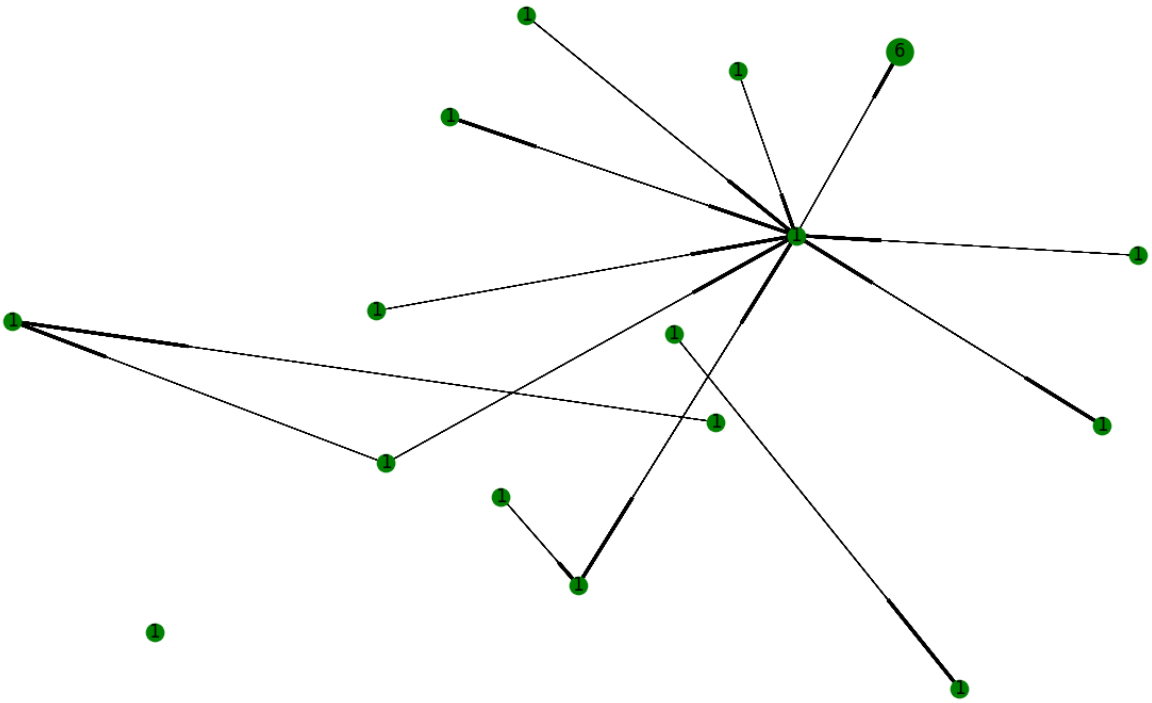
5. Shopping Dot

5.1. Condiciones

En el siguiente experimento se procesaran paquetes capturados en un shopping. La captura de los mismos se llevo a cabo a través de una red Wi-Fi que no tenia clave. La misma es ofrecida por el shopping para que los clientes puedan tener internet de manera gratis. No se esperan identificar muchos hosts, ya que durante la captura no había mucha gente

- Tamaño: Pequeña
- Horario: 5:30pm
- Duración: 20 minutos
- Modo: Wi-Fi

5.2. Grafo de la red



5.3. Porcentaje de protocolos

```
Porcentaje de paquetes broadcast 0.441

Probabilidades de protocolos

'IP':   97,9
'LLC':  1,06
'IPv6': 0,9
'ARP':  0,2
```

5.4. Información de los paquetes

```
Cantidad total de paquetes:     22229
Entropia:                       0.2011668780253044
Entropia Maxima:                2.584962500721156

symb: ('UNICAST',   'IP')       count: 21705    prob: 0.97642
symb: ('UNICAST',   'LLC')      count: 235      prob: 0.01057
symb: ('UNICAST',   'IPv6')     count: 190      prob: 0.00854
symb: ('BROADCAST', 'IP')       count: 58       prob: 0.00260
symb: ('BROADCAST', 'ARP')      count: 40       prob: 0.00179
symb: ('UNICAST',   'ARP')      count: 1        prob: 0.000004
```

5.5. Hosts distinguidos

```
average_info:       4.583774146463461

IP                  Info
'10.251.16.1'       1.1096244911744981
'10.251.16.240'     2.7725895038969277
```

5.6. Análisis

Como se esperaba, no se detectaron muchos hosts. Por otro lado se detectaron dos hosts distinguidos. Asumimos que el de menor información (mayor ocurrencia) es el router, pues debe saber constantemente quien esta conectado a cada IP.

Se puede observar que el porcentaje de paquetes broadcast es bajísimo. Esto se puede deber a la baja cantidad de host (y fija, pues no hubo mucho movimiento en el horario dado). Entonces, el uso intensivo de la red estuvo a cargo del router.

El símbolo mas probable es el <UNICAST, IP> lo cual tiene sentido , pues al tratarse de una red de un shopping, se espera que la red tenga un mayor consumo de datos (navegación por internet, vídeos, redes sociales)

Por ultimo, se ve que la entropía es muy baja (con respecto a la entropía máxima), lo cual tiene sentido, pues mayormente la red se utiliza para la transmición de datos.

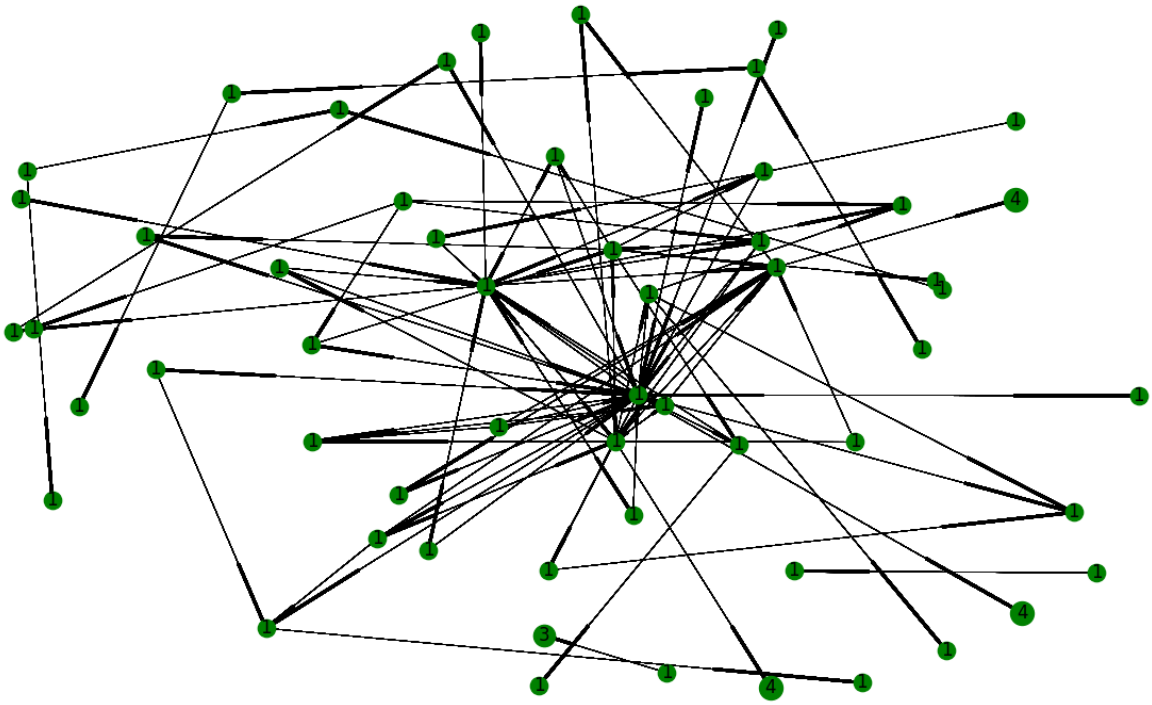
6. Experimento 3

6.1. Condiciones

Esta captura fue realizada en la oficina de trabajo de uno de los integrantes del grupo. Se trata de una oficina pequeña, en la que no hay demasiados dispositivos, por lo que la consideramos como una red privada de pequeño tamaño. Durante el período de captura utilizaban la red activamente unas cinco personas con sus respectivos dispositivos (PC y/o celular). Además, varias máquinas virtuales están conectadas a esta red.

- Tamaño: Pequeña
- Horario: 10 AM de un día laboral
- Duración: 126 minutos
- Modo: Wi-Fi

6.2. Grafo de la red



6.3. Porcentaje de protocolos

```
Porcentaje de paquetes broadcast 5.012

Porcentaje de aparición de cada protocolo

'ARP':   4,22
'IP':    95,4
'IPv6':  0,308
'EAPOL': 0,0026
```

6.4. Información de los paquetes

```
Cantidad total de paquetes: 149847
Entropia:                   0.37699955667490537
Entropia Maxima:            2.584962500721156

Simbolos de la fuente S1
Símbolo: ('BROADCAST', 'ARP')  Ocurrencias: 5981     Probabilidad: 0.03991
Símbolo: ('UNICAST', 'IP')     Ocurrencias: 141528   Probabilidad: 0.94448
Símbolo: ('BROADCAST', 'IP')   Ocurrencias: 1529     Probabilidad: 0.01020
Símbolo: ('UNICAST', 'IPv6')   Ocurrencias: 463      Probabilidad: 0.00308
Símbolo: ('UNICAST', 'ARP')    Ocurrencias: 342      Probabilidad: 0.00228
Símbolo: ('UNICAST', 'EAPOL')  Ocurrencias: 4        Probabilidad: 2.66938
```

6.5. Hosts distinguidos

```
Información promedio de los símbolos: 7.723940879124419

'10.0.0.19'       3.1146408504066163
'10.0.100.13'     5.571111068672806
'10.70.0.13'      4.8190385821163915
'200.69.224.105'  4.592970502636545
'10.0.0.190'      4.392773827414294
'10.70.0.141'     5.208540989288098
'10.152.199.140'  5.756028784590591
'10.0.0.235'      4.701581000568215
'10.1.1.66'       3.1959409525084643
'10.0.0.254'      5.793503490009254
'10.70.0.16'      3.701581000568215
'10.0.0.23'       3.0891751036353994
'10.0.0.213'      5.7437504548121545
'10.70.0.15'      5.793503490009254
'10.0.200.19'     6.087234693065964
'10.0.0.242'      5.845033790649336
'10.2.2.1'        5.695656166611109
'10.0.0.118'      4.264449730438754
'10.0.0.248'      5.719502908565477
'10.1.100.3'      6.056537895843048
'10.1.202.10'     6.041431003452839
'10.0.0.187'      5.486842151775202
```

6.6. Análisis

En primer lugar, observemos que para la fuente S1 el símbolo con más ocurrencias fue ('UNICAST', 'IP'). Creemos que esto puede deberse a que es una red pequeña, cuyos miembros no cambian constantemente como sí ocurre en una red pública. Entonces, no son necesarios muchos paquetes ARP para obtener las MAC address, ya que están cacheadas.

Por otra parte, para la fuente S2 obtuvimos muchos hosts distinguidos, y no hubo uno particular cuya información fuese notablemente menor que el resto, como sí ocurrió en otro experimento. Esto nos

permite concluir que por el hecho de ser una red pequeña, el router no envía una cantidad de paquetes extremadamente mayor que el resto de los hosts, y no pudo distinguirse entre el resto.

La entropía en este experimento, también se encuentra muy alejada de la entropía máxima. Notamos que es por la misma razón que en el experimento anterior.

7. Conclusion

Notamos diferencias entre las redes conectadas por WiFi y mediante cable. Estas diferencias, en realidad, pudieron darse por las particularidades de las red Ethernet (laboratorios del departamento de computación). Por ejemplo, la sincronización de datos entre las computadoras (y un servidor). Pudimos ver una mayor cantidad de tráfico broadcast en esta red, a causa de lo antes mencionado.

Acerca de la cantidad de muestras obtenidas en cada red, llegamos a la conclusión que, en redes grandes (por ejemplo, en el shopping Dot), al haber una poca cantidad de dispositivos conectados, la misma muestra no resulta representativa.

Otra manera posible para detectar símbolos distinguidos (es decir, los hosts de la red), utilizando otras herramientas a las usadas en la materia, sería analizar el tiempo de respuesta de los request. Es razonable pensar que, los request entre dos hosts que tienen un menor delay pertenecen a la misma red. Queda para un expermiento futuro, comprobar esto mismo, contrastandolo con los resultados obtenidos en este trabajo práctico.

En este trabajo práctico tuvimos la oportunidad de aplicar los conocimientos teóricos adquiridos durante las clases, sobre transmisión de información y sobre las capas 2 y 3 propuestas por el modelo OSI. A través de los diferentes experimentos, en distintos tipos de redes, pudimos percibir los diferentes comportamientos de los dispositivos de la red, en determinadas circunstancias. Además, los grafos de las redes nos permitieron observar más claramente qué es lo que estaba ocurriendo, para así encontrarle más sentido a los datos calculados por los scripts.