

Wiretapping de paquetes ARP: analizando redes locales usando Teoría de la Información

Gonzalo Ciruelos , Axel Maddonni y Federico Patané

Resumen—Con el objetivo de comprender diversos aspectos de una red local, nos planteamos analizar sus paquetes ARP usando herramientas de la teoría de la información. Para lograr eso, modelamos algún aspecto de nuestra red como una fuente de información, de manera de poder analizarla con las herramientas previamente dichas. Además, comparamos distintos tipos de redes para poder confirmar o refutar nuestras hipótesis generales.

Index Terms—ARP, LAN, broadcast, unicast, entropía, información, Ethernet, Wi-Fi

I. INTRODUCCIÓN TEÓRICA

EN este trabajo nos proponemos utilizar herramientas de la Teoría de la Información y los paquetes ARP para intentar comprender diferentes redes locales. La pregunta que nos intentaremos responder en este trabajo es ¿Podemos conocer la topología de la red y/o sus nodos más importantes?

La explicación de las fuentes utilizadas vendrá más adelante, pero es pertinente presentar el formato de los paquetes ARP. El protocolo de resolución de direcciones (Address Resolution Protocol) [1] es un protocolo usado para la resolución de direcciones de la capa de red a direcciones de la capa de enlace (usualmente MAC), lo que es una función crítica en redes de múltiple acceso.

Un paquete ARP tiene muchas partes, pero la que más nos interesa es el campo OPER, o sea el campo de Operación, que nos indica qué función cumple ese paquete. Los distintos valores de OPER en paquete ARP son is-at y who-has. Un paquete who-has es un paquete broadcast en el que una computadora pregunta al resto de la red quién es la que tiene una dirección de capa de red (generalmente IP) dada. La respuesta a ese paquete es un paquete is-at, que es un paquete unicast, enviado desde la computadora con la dirección de capa de red requerida hasta la computadora que emitió el who-has.

Por otro lado, con respecto a la teoría de la información, utilizaremos los conceptos de información y entropía clásicos.

II. DESARROLLO

COMO explicamos anteriormente, el trabajo se basará en analizar redes locales capturando paquetes ARP de redes locales y analizándolos utilizando herramientas de la teoría de la información.

Para capturar los paquetes de la red analizada, utilizamos el programa WIRESHARK. Luego, para postprocesar los datos y computar la información pedida utilizamos la

librería de Python SCAPY. SCAPY es una poderosa herramienta que permite capturar, decodificar, crear y enviar paquetes de manera muy sencilla.

A. Ejercicio 1

El ejercicio 1 proponía analizar los resultados modelando a los paquetes como una fuente de información muy simple: simplemente distinguir entre paquetes broadcast y paquetes unicast. La fuente consiste en dos símbolos, uno que representa a los paquetes unicast, y otro que representa a los paquetes broadcast.

Esta fuente no nos permitirá conocer muy bien quién es quién en la red, pero nos dará quizás un poco de información sobre lo que está pasando en la red local.

Que la cantidad de mensajes de broadcast sea muy grande, o sea, la información del símbolo que representa a los mensajes broadcast sea muy chica, será lo esperado. Esto se debe a que en una red normal, en la cual podemos escuchar todo lo que está pasando, lo esperable es que los dispositivos se comuniquen unos con otros en vez de estar enviando broadcasts todo el tiempo.

De alguna manera, si vemos pocos paquetes unicast, o sea que la información de este símbolo es muy alta, puede significar dos cosas:

1. No tenemos visibilidad total de la red, por ejemplo porque está switchheada, y entonces vemos solo los paquetes unicast dirigidos a nuestro host.
2. No hay comunicación efectiva entre los hosts de la red, porque los paquetes unicast de alguna manera miden cuanta comunicación de un host a otro está sucediendo.

Nuestra implementación de este ejercicio puede verse en el archivo `ejercicio1.py`.

B. Ejercicio 2

El ejercicio 2 proponía que diseñemos una fuente de memoria nula en base a los paquetes ARP, que nos permitiera distinguir a los nodos distinguidos de la red, para alguna definición que demos de nodo distinguido.

Experimentamos con varias fuentes, y terminamos seleccionando que la fuente sea el destino de los paquetes who-has, por las siguientes razones:

1. Si estamos analizando una red switchheada o muy subdividida en subredes virtuales (VLANs), entonces lo más probable es que solo recibamos paquetes who-has, dado que estos paquetes se envían en modo broadcast. Por esta razón los switches de la red local no los filtran y le llegan a todos los hosts. Quizás esta diferencia no sea notable en redes inalámbricas, pero en redes cableadas complejas, que generalmente tienen switches, puede hacer una gran diferencia.

2. Como justificamos anteriormente, vamos a usar paquetes who-has, ahora bien, la pregunta es porqué el destino de esos paquetes. En este caso la respuesta es más obvia: si un host es el destinatario de más paquetes ARP, entonces es más requerido por el resto de los hosts, entonces es más probable que sea un nodo distinguido, como por ejemplo un router.

Siguiendo estos simples preceptos, diseñamos nuestra fuente de información S1. Nuestra implementación de este ejercicio puede verse en el archivo `ejercicio2.py`.

C. Grafo de la red

Para los tres experimentos que realizamos, hicimos el grafo de la red que se desprende de los paquetes ARP enviados a lo largo de la captura.

El grafo fue realizado de la manera más natural. Por cada mensaje who-has capturado, el grafo tendrá una arista. Además, esa arista irá del nodo con IP igual a la IP fuente del who-has al nodo con IP destino del who-has.

Por razones de comodidad, solo mostramos aquellos nodos relevantes, es decir, los que tienen información baja con respecto a la fuente de información que S1.

Además, juntamos en uno a todos los hosts que tienen exactamente el mismo conjunto de aristas adyacentes. Esto se indicara con un [X]: si un nodo tiene [X] significa que ahí condensamos X cantidad de hosts, que tienen todos exactamente la misma conectividad que el nodo representado.

Además, marcamos con un cuadrado aquellos nodos distinguidos según la fuente S1, es decir, aquellos que tienen menos información que la entropía.

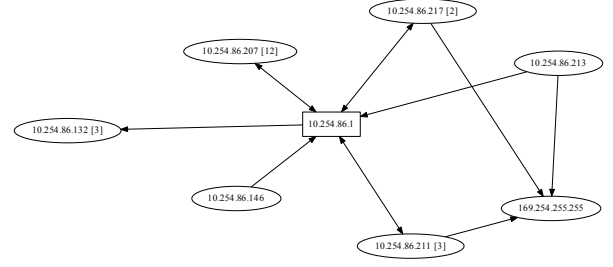


Figura 2

Grafo de conectividad de la red, inferido de los paquetes who-has. Para ver con mayor detalle, se puede hacer zoom-in en el pdf.

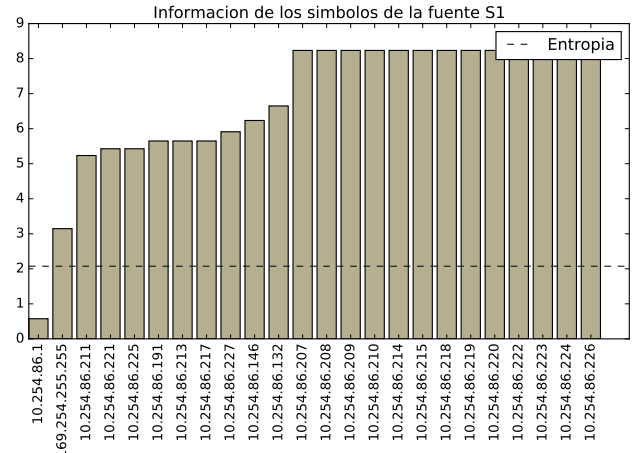


Figura 3

Información de los símbolos de la fuente S1: solamente los nodos con menor información son representados.

III. EXPERIMENTO 1

A. Resultados

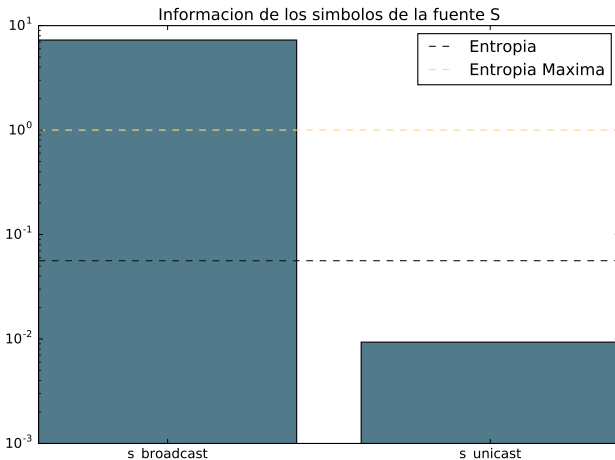


Figura 1

B. Discusión

IV. EXPERIMENTO 2

El segundo experimento que mostraremos fue realizado en los laboratorios del Departamento de Computación de FCEyN, más específicamente en el laboratorio 2. La captura de paquetes duró una hora y se realizó desenchufando el cable ethernet de la computadora 3 del laboratorio 2, y enchufándolo en nuestra propia computadora, de tal manera de poder capturar paquetes en modo promiscuo.

La captura de paquetes se realizó durante la tarde, mientras había aproximadamente 10 personas utilizando las computadoras de ese laboratorio. Las mediciones fueron realizadas con la autorización de uno de los administradores de la red, quien además ayudó a verificar las teorías que extrajimos de los paquetes. Veremos todo esto más adelante.

Antes de pasar a ver los resultados, una cosa que vale la pena decir es que, por como es la red de los laboratorios, al haber utilizado una computadora distinta de la de los labos para realizar las mediciones, la red no nos asignó IP,

es decir, no nos “conectó” de manera total a la red local. Sin embargo, esto no perjudicó en nada a las mediciones, y solo se vió reflejado en la fuente S.

A. Resultados

Primero analicemos la fuente S, que es la fuente de los paquetes unicast y broadcast. Como dijimos anteriormente, no teníamos dirección IP asignada. Esto sumado a que la red es switchcada, nos hace esperar que la cantidad de paquetes unicast que vamos a ver sea mucho más baja que lo normal. Esto, dicho en lenguaje de la Teoría de la Información, es que el símbolo que representa a los mensajes unicast tendrá una información más baja de la esperada.

Sin embargo, por como está configurada la red del laboratorio, sabemos que detrás de cada switch hay aproximadamente 4 o 5 computadoras, con lo cual podremos ver algunos paquetes unicast. Veamos qué pasó.

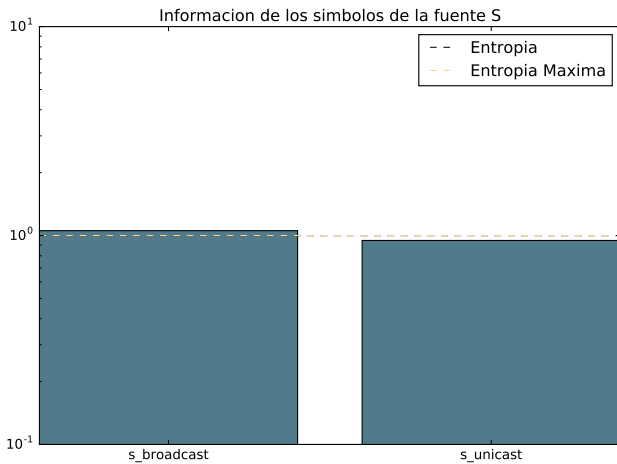


Figura 4

Ahora pasemos a ver el diagrama de conectividad de la red, que nos permitirá saber más sobre cómo está configurada.

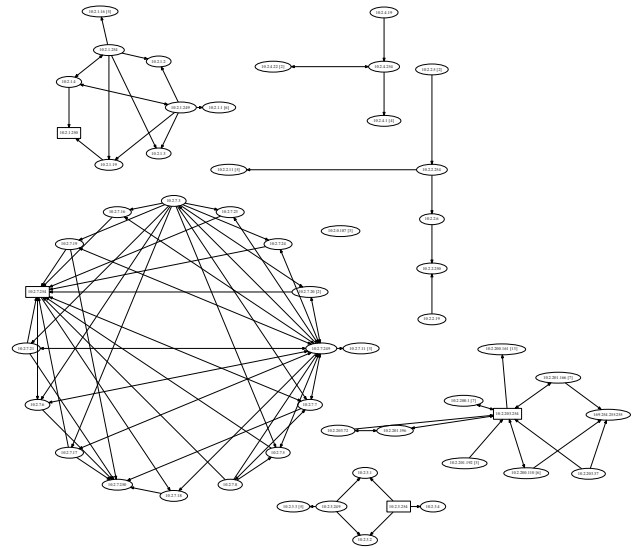


Figura 5

Grafo de conectividad de la red, inferido de los paquetes who-has. Para ver con mayor detalle, se puede hacer zoom-in en el pdf.

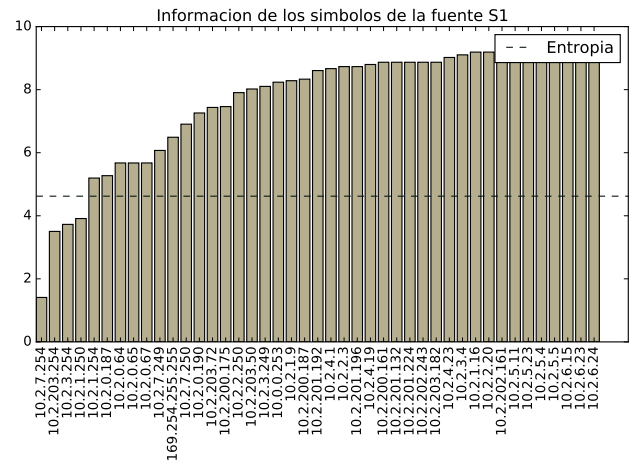


Figura 6

Información de los símbolos de la fuente S1: solamente los nodos con menor información son representados.

B. Discusión

Para empezar, se confirma lo que dijimos sobre la fuente S: la información del símbolo de los paquetes unicast es mucho más alta de lo esperada e iguala casi a la de los paquetes broadcast. Esto provoca que la entropía sea muy cercana a la máxima. Desde el punto de vista de la probabilidad, esta fuente es menos predecible que la que vimos antes.

Además, si no hubiéramos sabido como estaba la configurada la red (que se ve a simple vista en los laboratorios), podríamos haber inferido que está fuertemente switchcada dada la información de estos símbolos.

Con lo cual, como dijimos al principio del trabajo, esta fuente nos permite saber, entre otras cosas, que tanta visibilidad tenemos de toda la red: a más visibilidad más paquetes unicast nos van a llegar, por lo tanto más baja va a ser la información del símbolo asociado.

Para continuar con el grafo de conectividad y la fuente S1, como vemos, hay muchos dispositivos en la red, contamos más de 100 dispositivos. Sin embargo, solo mostramos aquellos relevantes para la discusión. Esta cantidad nos dice que la red es muy compleja, y que salvo excepciones, seguramente tendrá alguna subestructuración interna, por ejemplo subnetting, y que por lo tanto será un desafío intentar comprenderla.

Yendo más a lo específico, se ve que hay un grafo muy similar a un grafo completo, que tiene un nodo distinguido con IP 10.2.7.254. Nuestra hipótesis será que este dispositivo es el router. Además, todos los nodos de ese subgrafo tienen IPs de la forma 10.2.7.XXX. Es decir, ese subgrafo representa la subred 10.2.7.0/24. Por la cantidad y el detalle de nodos observados en esta subred, podemos inferir que es la red del laboratorio 2, donde capturamos los paquetes. Sin embargo de ninguna manera podemos imaginarnos por que tantas de las computadoras se envían paquetes ARP unas a otras. Veremos la razón de esto más adelante.

Además, podemos observar que se repite el mismo patrón en distintos lugares, por lo que podemos asumir que son los distintos laboratorios. El patrón al que nos referimos es la IP de la pinta 10.2.n.XXX, donde n es el número de laboratorio y XXX es el número de host. Además, como antes, 10.2.n.254 suele ser el nodo distinguido, con lo cual podemos asumir que todos esos son los routers de cada labo.

Hay otras 2 partes de la red que nos gustaría analizar. Primero, tres nodos de IP 10.2.0.XX, que lo único que hacen es mandar Gratuitous ARP. Como vimos anteriormente, un Gratuitous ARP son paquetes broadcast que actúan como paquetes is-at, pero que nunca fueron pedidos. Estos dispositivos actúan de manera anómala, pero como la red es muy compleja, podemos establecer como hipótesis que son dispositivos de control de la red.

Por último, podemos analizar la subred que se encuentra en la parte inferior izquierda de la figura. Como puede apreciarse, es bastante distinta al resto de las redes.

Esta red sigue un patrón muy similar al que observamos anteriormente en la red Wi-Fi. Además, aparece la misma IP que había aparecido antes: 169.254.255.255. Por último, esta subred tiene un nodo distinguido terminado en 254: 10.2.203.254, así que nuestra hipótesis será que esta subred representa parte de la red Wi-Fi de los laboratorios de la facultad.

Luego de plantear todas nuestras hipótesis, nos pareció interesante consultar con un administrador de la red para saber que tan exactas fueron nuestras predicciones utilizando un modelo tan simple como usamos.

Resultó que nuestras predicciones fueron muy acertadas. Como predijimos, cada labo se representa internamente con una subred. Estas subredes son las de la pinta 10.2.n.0/24. Además, los routers de cada labo son los de

IP 10.2.n.254, como habíamos predicho.

Finalmente, el subgrafo que creíamos que se correspondía con la red Wi-Fi, efectivamente lo hacía. Según el administrador, esa subred corresponde a IPs de un pool dinámico de IPs, que se asignan en el rango de direcciones de 10.2.200.0 a 10.2.201.255 de forma dinámica cada vez que un dispositivo nuevo se conecta a la red. Lo que el Wi-Fi de los laboratorios está configurado con un pool dinámico de direcciones, y 10.2.203.254 es el router de la red. Quizás era un poco difícil conocer toda esta información solamente con el grafo de conexiones, pero nos parece muy interesante contarlos de todas formas, porque es un patrón recurrente en redes Wi-Fi.

[poner porque las pcs se mandan paquetes unas a otras]
[poner los chabones de gratuitous ARP]

V. CONCLUSIÓN

REFERENCIAS

- [1] RFC 826 - Ethernet Address Resolution Protocol
<http://tools.ietf.org/html/rfc826>
C. Plummer 1982
- [2] <http://www.secdev.org/projects/scapy>