Teoría de las Comunicaciones: TP1

Federico Sabatini, Pedro Boitier, Dylan Socolobsky, Ignacio Losiggio

Introducción

En este trabajo realizaremos una captura de paquetes en distintos tipos de redes, analizando qué tipos de paquetes circulan en cada una y cómo esto se relaciona con las características de la misma (cantidad y tipo de dispositivos, medio de transmisión, topología, etc). Finalmente construiremos un pequeño análisis utilizando los conceptos de la teoría de la información vistos en la materia comentando cómo consideramos que las características propias de la red en cuestión se ven reflejadas en el análisis de la red como fuente de información.

Métodos y condiciones de los experimentos

Para la captura de paquetes utilizamos la biblioteca Scapy del lenguaje Python. Ésta nos permite capturar paquetes y analizar su contenido.

Por cada red capturamos 20000 paquetes esperando que ésta sea una muestra significativa de su comportamiento. Es importante acotar que ninguna de las redes capturadas tiene topología de bus, por lo que los paquetes recibidos tenían cómo destinatario al computador en el cual se realizó la captura.

Para trabajar la red con los conceptos de la teoría de la información, modelamos la red como una fuente de información de memoria nula. En esta, cada tipo de paquete es un símbolo distinto. En el programa adjunto (network_scapy_ej1.py) se calcula la información de cada cada tipo de paquete y la entropía de la red según el modelo propuesto.

Redes analizadas

Dylan Socolobsky	**Tipo de tecnología:** Ethernet y WiFi (captura realizada en ethernet)
	Dispositivos: 3 (computadora de escritorio, celular y router)
Ignacio Losiggio	**Tipo de tecnología:** Ethernet y WiFi (captura realizada en ethernet)
	Dispositivos: 15 (2 televisores, 3 pcs de escritorio, 4 celulares, 2 routers, 3 notebooks, 1 CCTV)
	Horario de la captura: 19:55hs de un viernes
Federico Sabatini	**Tipo de tecnología:** Ethernet y WiFi (captura realizada en ethernet)
	Dispositivos: 8 (4 computadoras de escritorio, 3 celulares y router).
	Nota: La captura fue realizada desde una VM de Virtualbox. Cuya placa de red virtualizada se configuró con la opción Bridged-Networking[1]. Que permite recibir los broadcasts de la red.
Pedro Boitier	**Tipo de tecnología:** Ethernet
	Dispositivos: 2 (computadora de escritorio y router)
	Horario de la captura: entre las 20:38 y 20:42 de un viernes

Resultados de los experimentos

Resultados Pedro Boitier

Ethertype	**Símbolo**	**Probabilidad**	**Información**
2048	UNICAST-IP	0.82100	0.28455
34525	UNICAST-IPv6	0.17820	2.48843
2054	UNICAST-ARP	0.00060	10.70275
35130	UNICAST-IEEE 1905.1	0.00020	12.28771
Entropía de la fuente: 0.68593			

[1] Manual de virtualbox. https://www.virtualbox.org/manual/ch06.html#network_bridged

Resultados Ignacio Losiggio

Ethertype	**Símbolo**	**Probabilidad**	**Información**
2048	UNICAST-IP	0.97455	0.03719
2048	BROADCAST-IP	0.00535	7.54625
34525	UNICAST-IPv6	0.17820	5.97937
2054	BROADCAST-ARP	0.00155	9.33352
2054	UNICAST-ARP	0.00020	8.53282
Entropía de la fuente: 0.20890			

Resultados Federico Sabatini

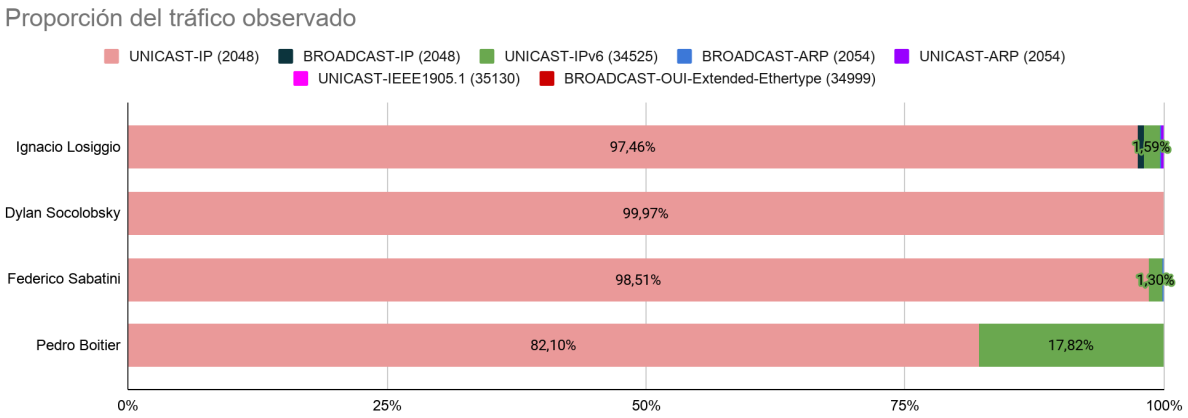
Ethertype	**Símbolo**	**Probabilidad**	**Información**
2048	UNICAST-IP	0.98505	0.02173
34525	UNICAST-IPv6	0.01295	10.58727
2054	UNICAST-ARP	0.00065	6.27090
35130	BROADCAST-ARP	0.00115	9.76415
34999	BROADCAST-OUI extended ethertype	0.00020	12.28771
Entropía de la fuente: 0.12318			

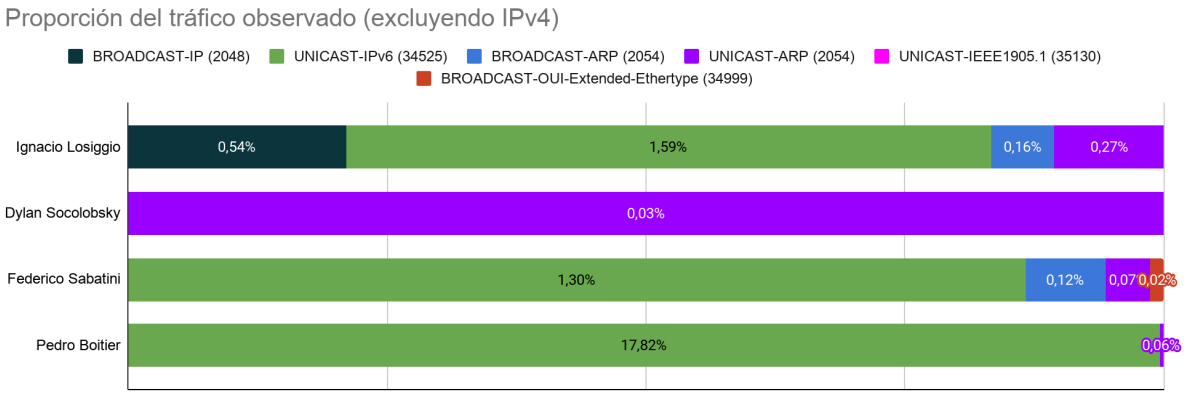
Resultados Dylan Socolobsky

Ethertype	**Símbolo**	**Probabilidad**	**Información**
2048	UNICAST-IP	0.99970	0.00043
2054	UNICAST-ARP	0.00030	11.70275
Entropía de la fuente: 0.00394			

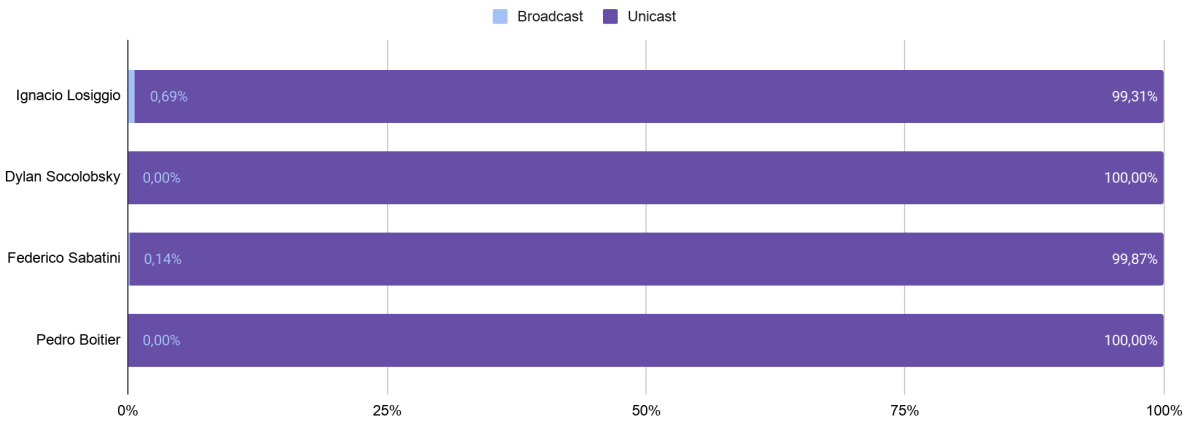
Resumen de la información capturada

Porcentaje de aparición de cada protocolo





Porcentaje de broadcast contra unicast



En las redes de Dylan y Pedro no se capturaron paquetes broadcast, suponemos que por el pequeño tamaño de las mismas.

Llama la atención la gran proporción de paquetes que son Unicast IPv6 en la red de Pedro. Ninguna otra red está cerca de esa proporción. Suponemos que se debe a que el router de esa red es más nuevo que los demás y por tanto utiliza protocolos más nuevos dónde es posible.

En las redes de Federico y Pedro encontramos tráfico que no esperábamos e intentaremos describir en la siguiente sección.

Conclusiones

¿Considera que las muestras obtenidas analizadas son representativas del comportamiento general de la red?

Dado que las capturas se dieron fuera del horario laboral creemos que el tráfico observado es representativo de la red durante el tiempo “ocioso” de sus usuarios. Pese a esto no creemos que las proporciones de los protocolos cambien mucho en otra franja horaria, dado que gran parte del tráfico es software actualizándose, clientes de mensajería haciendo “polling” de mensajes nuevos, etc.

¿Hay alguna relación entre la entropía de las redes y alguna característica de las mismas (ej.: tamaño, tecnología, etc)?

La única relación que encontramos es que ante más variación de los dispositivos conectados a la red hay mayor es la diversidad de protocolos que observamos. Aún así la mayor entropía la tiene la red de Pedro (que es relativamente modesta, con sólo dos dispositivos), dado que posee una cantidad considerable de tráfico IPv6.

Por esto no creemos que nuestra suposición se mantenga firme con los datos encontrados. Para poder refinarla podríamos tener en cuenta también la fecha de fabricación de los dispositivos de red. Suponemos que dispositivos muy recientes administran una red IPv6 además de una IPv4.

¿Considera significativa la cantidad de tráfico broadcast sobre el tráfico total?

Es insignificante comparada con el total. En las redes analizadas el tráfico broadcast es ARP en su mayoría. La excepción en la red de Ignacio es causada por protocolo de descubrimiento de hosts (NetBIOS) y en la de Federico creemos que es algún paquete de control de su módem.

¿Cuál es la función de cada uno de los protocolos encontrados? ¿Cuáles son de control? ¿Cuáles transportan datos del usuario?

Protocolo	**Función**	**Tipo**
IP	Transporte de paquetes de información	Datos
IPv6	Transporte de paquetes de información pero bajo IPv6	Datos
ARP	Descubrimiento de dispositivos físicos en la red	Control
IEEE 1905.1	Capa de abstracción sobre varias tecnologías de comunicación de red domésticas tales cómo ethernet, coaxil o wireless. Permite el autodescubrimiento de la topología de la red.	Control
OUI Extended Ethertype	OUI (Organizationally unique identifier) es un mecanismo que permite a una organización asignar identificadores únicos a sus propios protocolos.	Control

¿En alguna red la entropía de la fuente alcanza la entropía máxima teórica?

La entropía máxima teórica está dada por la fórmula $H_{max}(S) = log_2 n$ siendo n la cantidad de símbolos de la fuente. Dicho esto tenemos varias opciones:

1. Asumir que los protocolos observados en la captura de cada red son todos los protocolos existentes para esa red (n depende de la red).
    - Federico Sabatini: $n = 5$, $H_{max}(S) \approx 2.321$
    - Dylan Socolobsky: $n = 2$, $H_{max}(S) = 1$
    - Pedro Boitier: $n = 4$, $H_{max}(S) = 2$
    - Ignacio Losiggio: $n = 5$, $H_{max}(S) \approx 2.321$
2. Asumir que los protocolos observados entre todas las capturas son todos los protocolos existentes para esas redes ($n = 7$, $H_{max}(S) \approx 2.807$).
3. Asumir que todos los ethertypes posibles (16bits) pueden transmitir tanto por broadcast cómo por multicast ($n = 2^{16+1}$, $H_{max}(S) = 17$).

Dicho esto, ninguna de las redes se acerca a los valores teóricos.

¿Ha encontrado protocolos no esperados? ¿Puede describirlos?

- OUI Extended Ethertype: El cual se estima que es emitido por el **módem del ISP** (Speedy ADSL) pues al analizar los paquetes con wireshark, se identifican como Quantenna Communications, Inc. (empresa con la cual trabaja Telefónica), y son emitidos por un **MitraStar 98**, un router que poseemos cuyo distribuidor es Movistar.

  Estimamos que el router utiliza este protocolo para control interno pues OUI (organizationally unique identifier) se utiliza por los proveedores para el desarrollo de protocolos específicos propios **pero desconocemos su uso real**.

- IEEE1905.1: Emitido por el módem de la red el módem permite la autoconfiguración de la misma (por ejemplo facilitando la implementación de balanceo de carga en redes híbridas).

Observación del tráfico ARP para distinción de hosts

A continuación decidimos extender el modelo enfocándonos solo en los paquetes ARP. Para ello decidimos realizar una nueva captura en la red de Ignacio Losiggio, su red fue elegida por ser la que posee un mayor tráfico de este tipo.

La captura en cuestión posee 1000 paquetes ARP y fue realizada con el software Wireshark. Usamos scripts en python con la biblioteca scapy para procesar los datos una vez capturados. La captura fue realizada el 3 de Mayo a las 10 PM.

Como primera observación buscaremos identificar cuáles son los nodos de la red. Para ello contaremos sus apariciones como fuente y/o destino. Hecho esto realizaremos el mismo análisis que hicimos para las capturas anteriores.

Un detalle importante de este modelo es que le asigna dos símbolos a cada paquete ARP, haciendo que los announcements “cuenten doble”.

Resultados obtenidos

n°	IP address	#src	#dst	Probabilidad	Información
1	192.168.1.1	333	603	0.42740	1.22635
2	192.168.1.177	458	0	0.20913	2.25751
4	192.168.1.190	138	138	0.12603	2.98819
3	192.168.1.106	130	130	0.11872	3.07435
5	192.168.1.187	1	198	0.09087	3.46009
6	192.168.1.102	14	1	0.00685	7.18982
7	192.168.1.240	4	5	0.00411	7.92679

7	192.168.1.113	9	0	0.00411	7.92679
7	192.168.1.246	0	9	0.00411	7.92679
10	192.168.1.162	3	3	0.00274	8.51175
11	192.168.1.104	4	3	0.00320	8.28936
12	192.168.1.175	1	3	0.00183	9.09672
13	192.168.1.101	0	2	0.00091	10.09672
Entropía de la fuente: 0.21600					

Conclusiones

¿La entropía de la fuente es máxima? ¿Qué sugiere esto acerca de la red?

Supongamos que la red consta sólo de las IPs observadas. Si todos los eventos que ocurrieron hubieran sido equiprobables entonces la entropía sería máxima. Informalmente podemos observar que esto no ocurre sin siquiera realizar las cuentas correspondientes.

Si tomamos nuevamente $H_{max}(S) = log_2 n$ asumiendo $n = 13$ la cantidad total de hosts de la red luego $H_{max}(S) \approx 3.7004397$.

La entropía de la red es baja $H(S) \approx 0.21600$. Lo cual tiene sentido pues la mayoría del tráfico ARP son broadcast *"Who has 192.168.1.1?".*

Realizando una inspección ocular con wireshark podemos localizar la fuente de este tráfico:

No.	Time	Source	Destination	Protocol	Length	Info
38	36.175726239	SamsungE_b7:5c:8b	Broadcast	ARP	60	Who has 192.168.1.1? Tell 192.168.1.177
41	39.191686291	SamsungE_b7:5c:8b	Broadcast	ARP	60	Who has 192.168.1.1? Tell 192.168.1.177
42	42.207570446	SamsungE_b7:5c:8b	Broadcast	ARP	60	Who has 192.168.1.1? Tell 192.168.1.177
45	45.223364964	SamsungE_b7:5c:8b	Broadcast	ARP	60	Who has 192.168.1.1? Tell 192.168.1.177
47	48.239317382	SamsungE_b7:5c:8b	Broadcast	ARP	60	Who has 192.168.1.1? Tell 192.168.1.177
50	51.255247754	SamsungE_b7:5c:8b	Broadcast	ARP	60	Who has 192.168.1.1? Tell 192.168.1.177
51	54.271210748	SamsungE_b7:5c:8b	Broadcast	ARP	60	Who has 192.168.1.1? Tell 192.168.1.177

Dicho mensaje se repite cada aproximadamente 3 segundos por un host específico. Concluimos entonces no solo que la entropía **no es máxima**, sino que además la red posee tráfico altamente predecible.

¿Se pueden distinguir nodos? ¿Se les puede adjudicar alguna función específica?

Nuestra suposición original fué que el dispositivo con más *"Who has …?"* iba a ser probablemente el gateway de la red. Sumado a esto encontramos otros cuatro dispositivos con comportamiento distinguido.

¿Hay evidencia parcial que sugiera que algún nodo funciona de forma anómala y/o no esperada?

Podemos dividir los hosts observados en cinco grupos:

- 192.168.1.1: El host que más veces apareció en total (y por tanto el que menos información aporta).
- 192.168.1.177: El segundo host que más veces apareció. Todas sus apariciones son cómo src, lo cual resulta extraño.
- 192.168.1.190, 192.168.1.106: Hosts que únicamente aparecieron por sus propios announcements. Ocupan el tercer y cuarto puesto en menor cantidad de información.
- 192.168.1.187: El quinto host en menor cantidad de información, casi todas sus apariciones son cómo parte de un “Who has?”
- El resto de los hosts, si bien hay variación en la cantidad de símbolos observados (15 el que más tiene y 2 el que menos) creemos que está opacado por el comportamiento de los hosts antes mencionados.

¿Existe una correspondencia entre lo que se conoce de la red y los nodos distinguidos detectados por la herramienta?

- 192.168.1.1: El router que posee la salida a internet. Creíamos que iba a ser el más solicitado y así resultó una vez obtuvimos la captura.
- 192.168.1.177: Un SmartTV. Entendíamos que la calidad del software que corre en estos dispositivos no era alta, pese a eso no podemos comprender en su totalidad la razón del comportamiento observado.
- 192.168.1.190, 192.168.1.106:
    - 190 es un dispositivo para la administración de cámaras de vigilancia. Su IP está asignada estáticamente y creemos que es por eso que se anunció tan seguido.
    - 106 es la notebook dónde se realizó el scan, por ésto vemos la totalidad de sus announcements.
    - Suponemos que los announcements de IPs administradas por el router están siendo filtrados por el router mismo para prevenir un ataque de ARP poisoning. Esto explicaría por qué todos los announcements de la captura provienen de una IP no asignada vía DHCP.
- 192.168.1.187: **No sabemos qué dispositivo es.** Durante la escritura del informe revisamos la tabla ARP del router y las IPs asignadas por DHCP. Ni la IP ni la MAC

correspondiente figuraban en éstas[2]. Adicionalmente, la MAC que aparece en la captura no figura en las bases de fabricantes que consultamos.

[2] Esto es lo que en la jerga técnica se lo conoce cómo *“un misterio”*.