TP1: Wiretapping

Teoría de las Comunicaciones
Departamento de Computación
FCEN - UBA

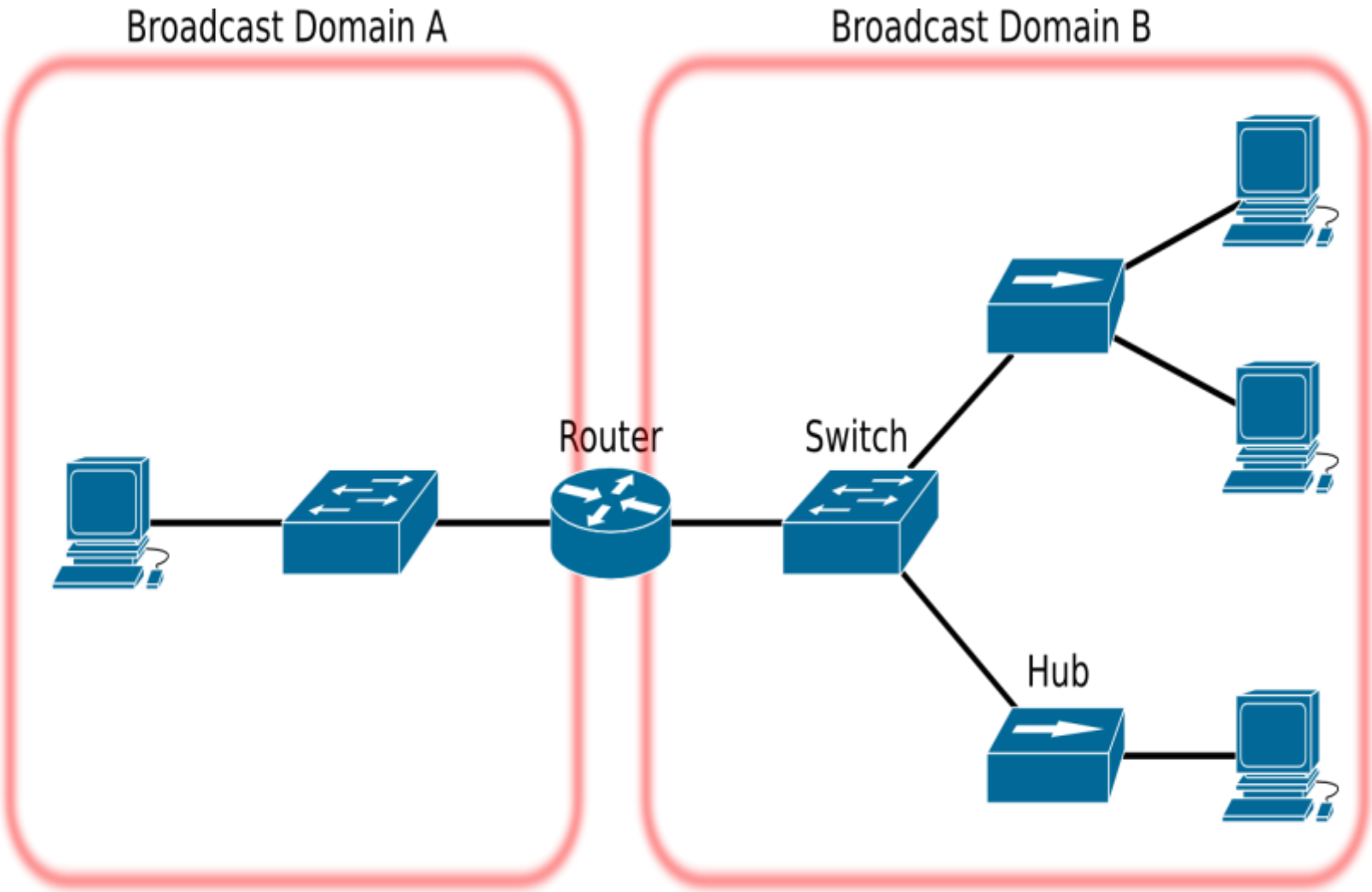
Objetivos

- Utilizar técnicas provistas por la **teoría de la información** para estudiar diversos **aspectos** de una **red** de manera analítica.
- Utilizar **herramientas** de manipulación y análisis de paquetes frecuentemente usadas en el dominio de las **redes** de computadoras.
- **Experimentar** con **redes**. No siempre es lo que parece.
- Enmarcar el análisis en un informe con estructura de **informe científico** (introducción; métodos y condiciones de los experimentos; resultados de cada experimento; conclusiones).
- El código no es tan importante. Ojo con las figuras. Que sean claras y tengan leyendas.

Capturando tráfico

- Wireshark, Scapy.
- Grupos: 3 o 4 integrantes.
- Una captura por cada miembro del grupo.
- Capturar > 10.000 tramas.
- Cada captura en una **red distinta.**
- Al menos una de las capturas debe ser sobre una red mediana/grande no controlada (laboratorios, red de trabajo, shopping, etc.).
- Debe haber capturas tanto desde enlaces cableados como inalámbricos.

Capturando tráfico



Capturando tráfico

Trama Ethernet

La forma en que las redes Ethernet trasmiten sus datos se llama **trama.** Las tramas tienen una longitud mínima de 64 bytes y una longitud máxima de 1518 bytes (con 802.1Q un total de 1522).

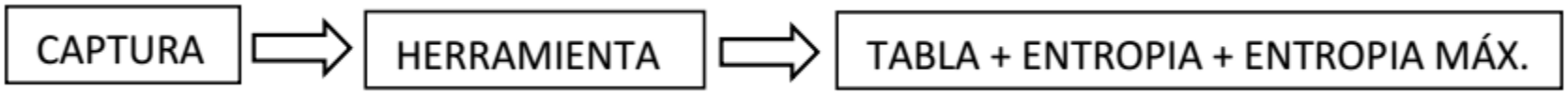
FORMATO DE TRAMA

Preámbulo	Delimitador de inicio de trama	MAC de destino	MAC de origen	Etiqueta 802.1Q (opcional)	Longitud /Tipo	Encabezado y datos	Secuencia de verificación de trama (FCS)
7 Bytes	1 Bytes	6 Bytes	6 Bytes	4 Bytes	2 Bytes	46 a 15000 Bytes	4 Bytes

Modelando tráfico: Fuente S1

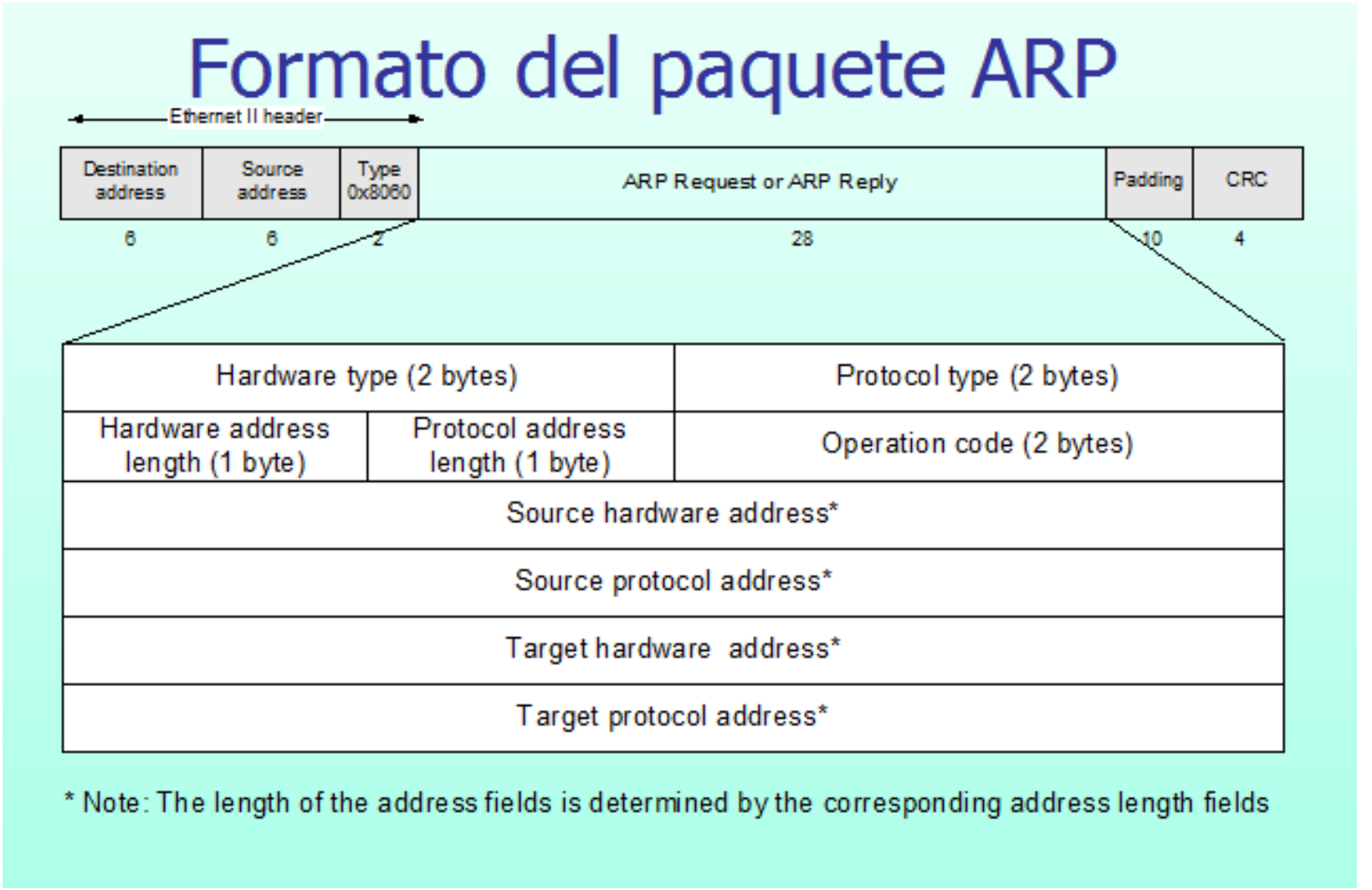
- Sean p1...pn las tramas de capa2 que se capturan en una red local. Se pueden modelar las tramas capturadas como una fuente de información de memoria nula **S1 = {s_1 ,s_2 ,...,s_q }**, donde cada s_i está formado por la combinación entre el tipo de destino de la trama (unicast o broadcast) y el protocolo de la capa inmediata superior encapsulado en la misma.
- **Por ejemplo, s_i =<broadcast, ARP>**
- Implementar una herramienta usando Scapy, que capture tráfico en una red local y muestre representativamente la fuente modelada S1.
- La salida debe consistir en una tabla que muestre la probabilidad e información de cada símbolo, la entropía de la fuente y su entropía máxima.
- http://www.secdev.org/projects/scapy
  - *"Scapy is a powerful interactive packet manipulation program. It is able to forge or decode packets of a wide number of protocols, send them on the wire, capture them, match requests and replies, and much more."*

Modelando tráfico: Fuente S2

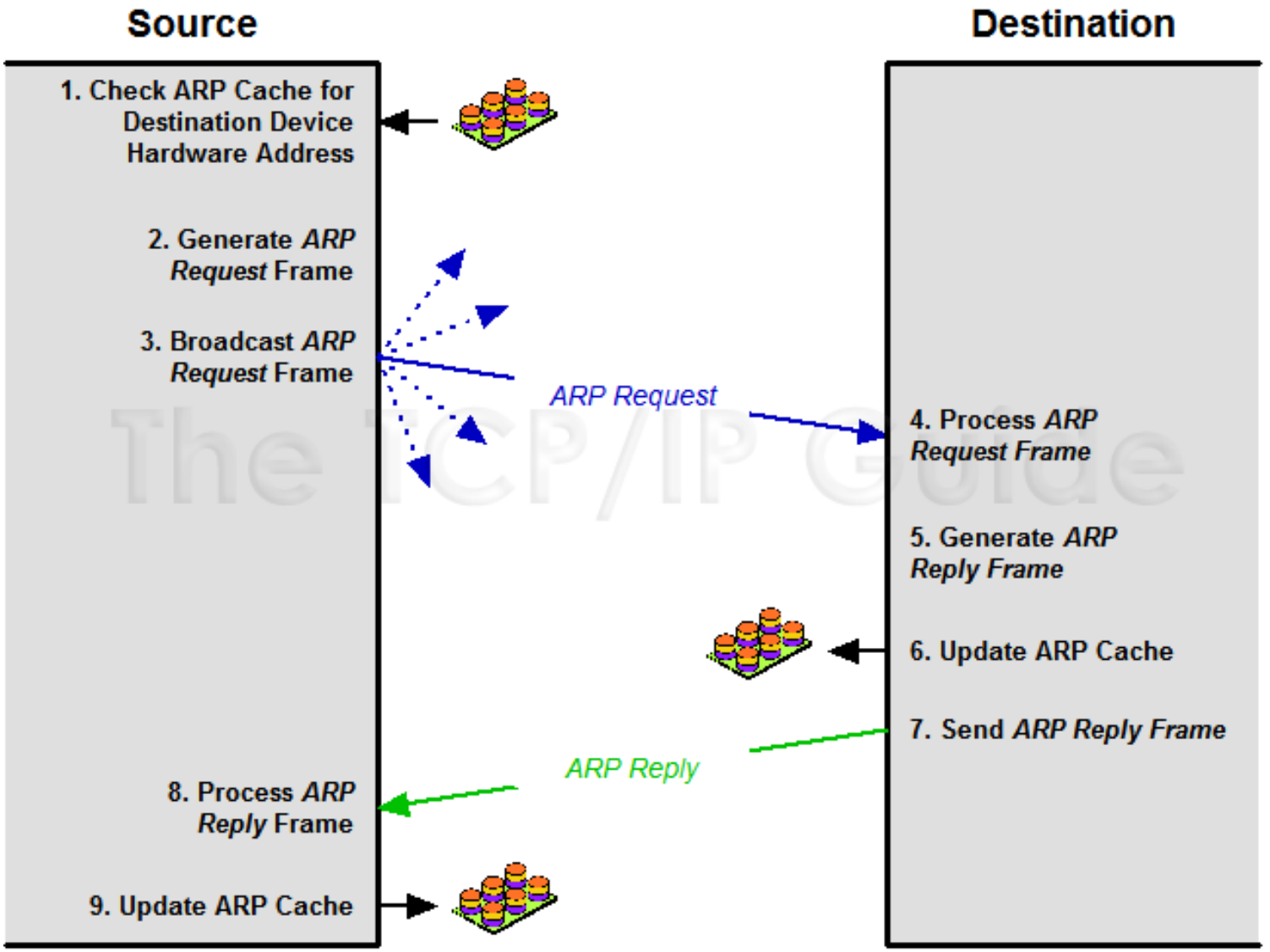


- Proponga un modelo de fuente de información de memoria nula **S2** con el objeto de distinguir los hosts de cada red.
- La distinción de **S2** debe estar basada únicamente en las **direcciones IP dentro de los paquetes ARP**.
- El criterio para el modelado lo deberá establecer cada grupo utilizando las herramientas teóricas provistas por la teoría de la información.
- Se puede pensar que un símbolo es **distinguido** cuando sobresale del resto en términos de la información que provee.
- Adapte o extienda la herramienta anterior para que funcione con la nueva fuente.

Modelando tráfico: Fuente S2



Modelando tráfico: Fuente S2



Address Resolution Protocol (ARP) Transaction Process

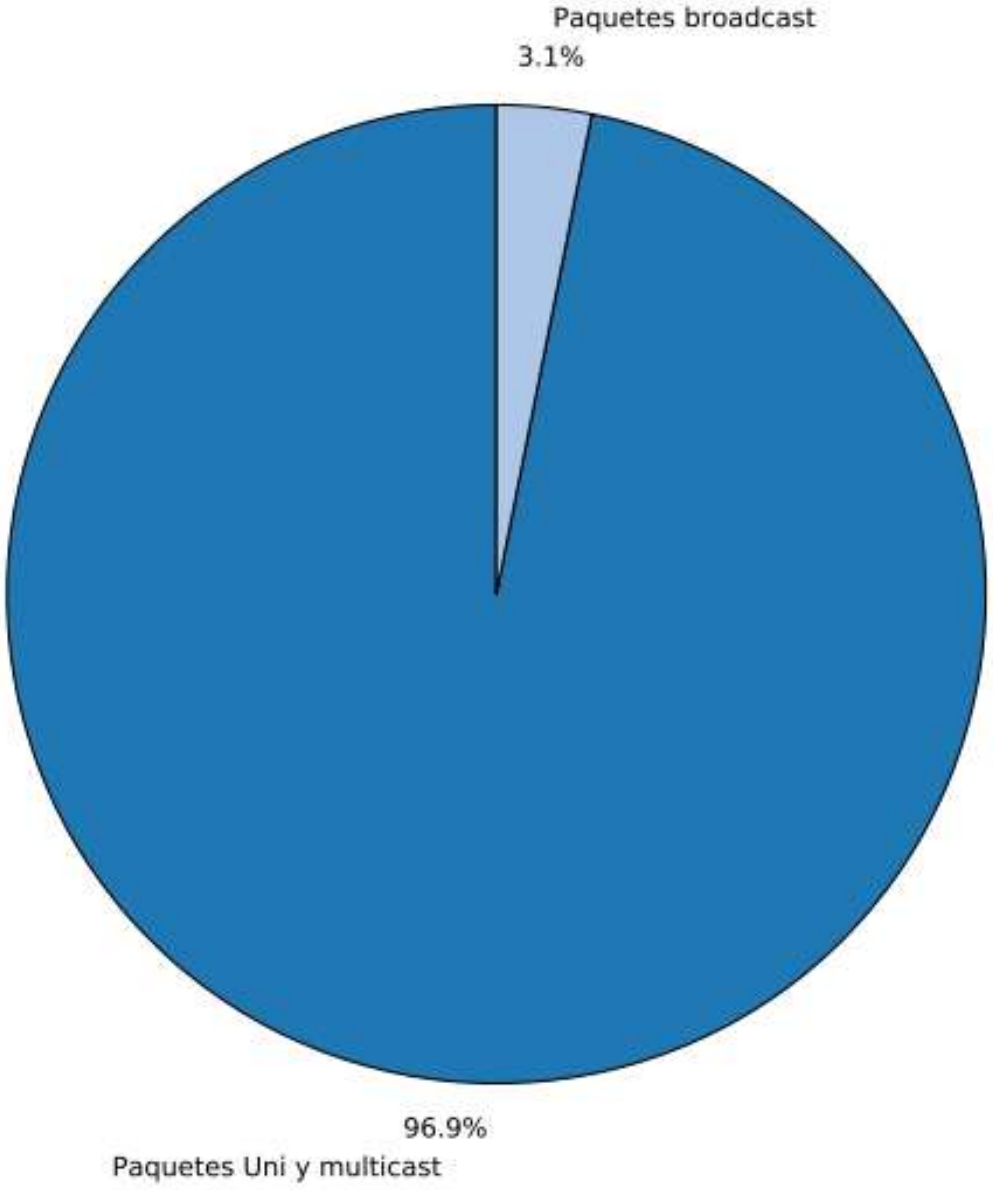
Gráficos y análisis

- Utilizando estas herramientas realizar experimentos para analizar los **símbolos distinguidos** en cada una de las fuentes. Los análisis deben estar basados en conceptos formales de la **teoría de la información**.
- Específicamente, se debe analizar qué símbolos se puede distinguir en cada red, observando la **diferencia entre su información y la entropía de la fuente.**
- El informe debe seguir la siguiente estructura: i) introducción; ii) métodos y condiciones de los experimentos; iii) resultados de cada experimento; iv) conclusiones.
- Formatting IEEE Papers http://mocha-java.uccs.edu/ieee/
- Entre los métodos y condiciones de cada experimento se debe detallar la descripción de la red -tipo, tamaño, modo de acceso, etc.-, las características de la muestra -tamaño, horario, día de la semana, etc.- y la justificación de la elección del modelo de la fuente S2.

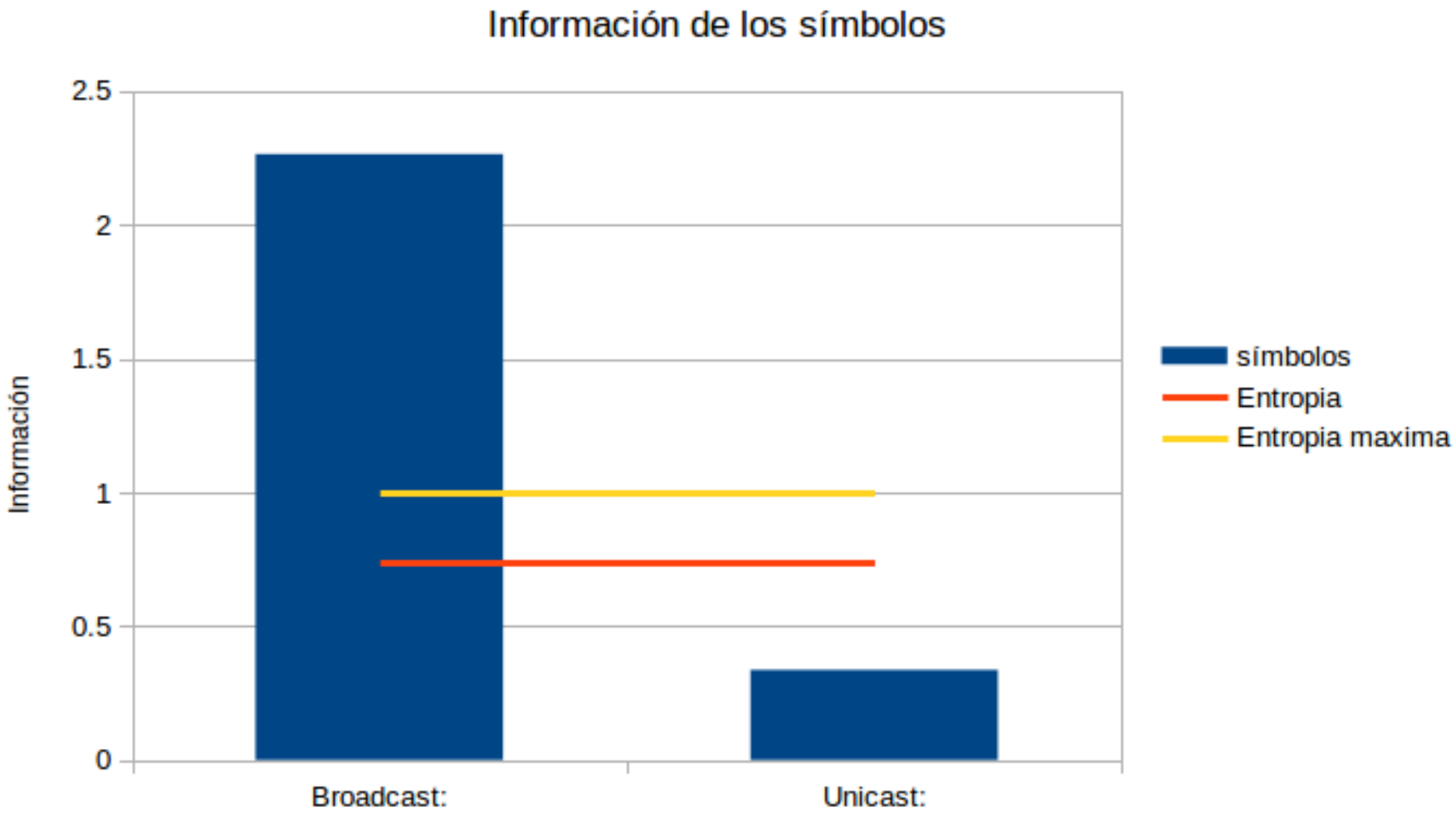
Gráficos y análisis

- La presentación de los resultados debe efectuarse **para cada red** mediante, al menos, los **gráficos** sugeridos a continuación:
- Para S1: Mostrar la cantidad de información de cada símbolo comparando con la entropía de la fuente y la entropía máxima. Mostrar el porcentaje de tráfico broadcast sobre el tráfico total. Mostrar el porcentaje de aparición de cada protocolo encontrado.
- Para S2: Mostrar la cantidad de información de cada símbolo comparando con la entropía de la fuente y la entropía máxima. Dados los paquetes ARP, mostrar mediante un grafo, la red de mensajes ARP subyacente (de ser necesario, agrupar adecuadamente varios nodos en uno para mejorar la visualización).

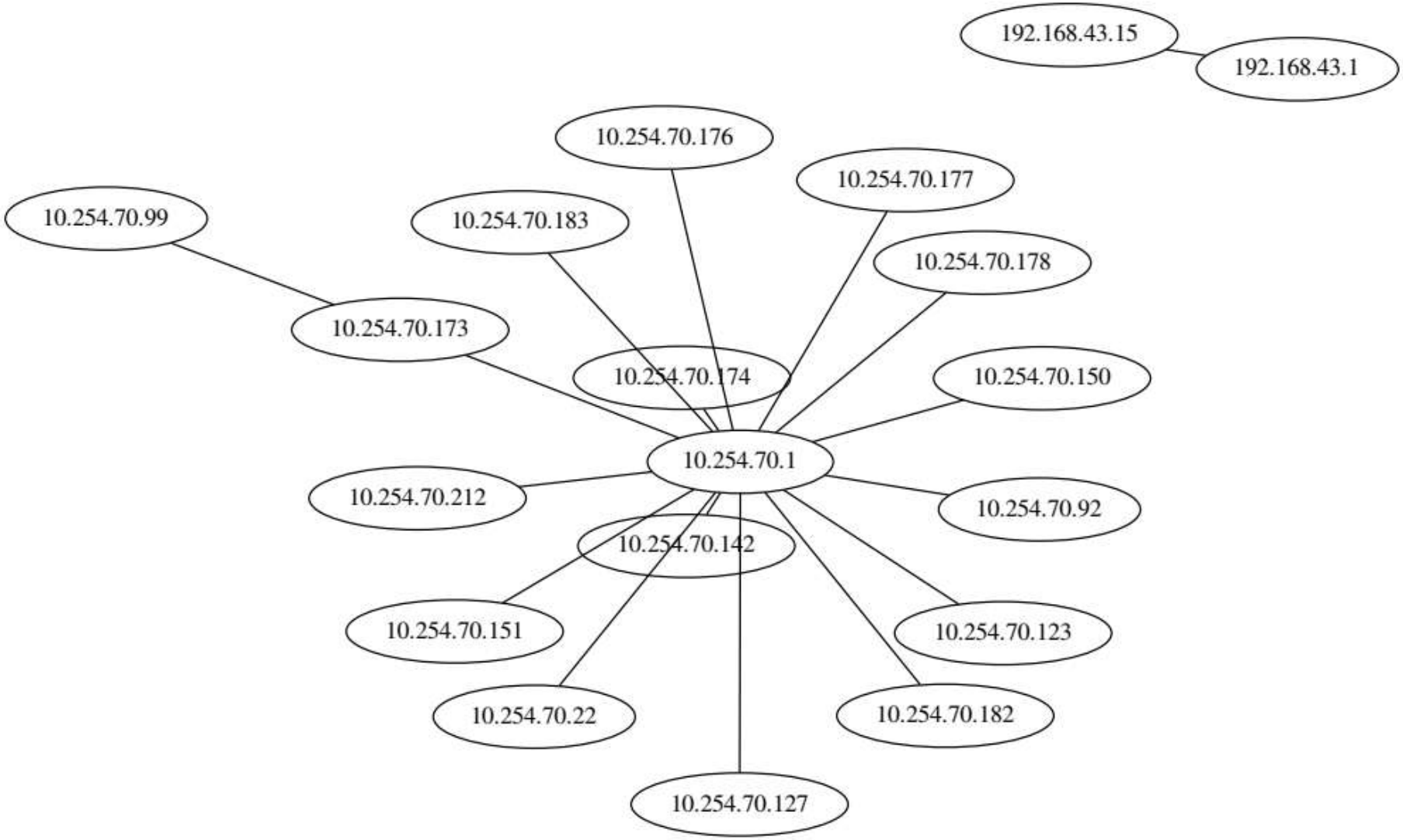
Porcentaje de tráfico broadcast sobre el tráfico total



Cantidad de información de cada símbolo comparando con la entropía de la fuente



Red de mensajes ARP subyacente



Análisis para cada red

- A su vez los resultados por experimento deben responder **para cada red**, algunas de las preguntas descriptas a continuación.
- No hace falta transcribirlas en el informe y se valorará significativamente el planteo de nuevas preguntas.
  - Para S1: ¿Considera significativa la cantidad de tráfico broadcast sobre el tráfico total? ¿Cuál es la función de cada uno de los protocolos encontrados?, etc.
  - Para S2: ¿La entropía de la fuente es máxima? ¿Qué sugiere esto acerca de la red? ¿Bajo qué condiciones la entropía sería máxima? ¿Se pueden distinguir nodos? ¿Se les puede adjudicar alguna función específica?, etc.

Análisis global

- A continuación, se sugieren preguntas para responder a la hora de realizar un análisis global en la conclusiones.
- Se valorará significativamente el planteo de nuevas preguntas.
  - De haber diferentes tamaños de redes, ¿Aprecia alguna diferencia desde el punto de vista de las fuentes de información analizadas?
  - ¿Ha notado alguna diferencia durante la captura de datos entre el acceso a la red mediante WiFi y el acceso mediante cable? ¿A qué se lo atribuye?
  - Etc.

¿PREGUNTAS?