DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Taller 1

Wiretapping

Teoria de las comunicaciones

Segundo Cuatrimestre de 2020

Integrantes	LU	Correo electrónico
Luis Castro	422/14	castroluis1694@gmail.com
Felipe Mateo Curti	71/17	fmcurti@gmail.com
J. Martin Mamani Aleman	630/17	mr.tinchazo@gmail.com
Ian Luca Celestino Breitman	375/15	ian13celestino@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. Introduction

En la era de la información (o desinformación) vivimos en un constante intercambio de datos. Esto, por supuesto, se ve mejor reflejado en nuestro estado permanente de conexión a Internet. Es este flujo inacabable de bytes sobre el que estaremos trabajando en este taller.

Analizaremos el tráfico de 4 redes domésticas por medio de un script de Python y la biblioteca Scapy provista por la cátedra. Recolectaremos información sobre los paquetes recibidos mediantes sniffing y compararemos los datos encontrados para cada red utilizando varios criterios: el tipo de transmisión (Unicast y Broadcast), la entropía y la proporción de protocolos en base al total de paquetes analizados.

Para la detección de paquetes deberemos modificar el script de Python provisto por la cátedra para hacer los cálculos pertinentes y posteriormente mostrar la información obtenida con algún recurso gráfico.

2. Métodos y condiciones de los experimentos

Para estos experimentos modificamos el script dado por la cátedra, ahora este no solo calcula la entropía de la fuente sino que también almacena todos los datos relevantes para la experimentación en un csv, ya sea el porcentaje de Unicast/Broadcast como los distintos tipos de protocolos que encontró.

Para los experimentos utilizamos 4 tipos de redes distintas que poseen estas características:

- **Red1:** Esta es una red Wifi de una familia compuesta por 3 personas, esta red es usada normalmente para ver videos/navegar. Esta red posee alrededor de 7 dispositivos de los cuales aunque sea 4 estan siendo usado de forma continua.
- **Red2:** Otra red WiFi de uso familiar, usada normalmente por 3 celulares y 3 computadoras de escritorio.
- **Red3:** Esta es otra red de uso familiar. Posee aproximadamente 10 dispositivos de los cuales 3 se utilizan continuamente y cuenta con ethernet IEEE 802.3 y wifi Realtek RTL8723BE 802.11n.
- **Red4:** Red usada por una única persona con computadora de escritorio conectada por cable Ethernet, computadora de uso laboral conectada por Wi-Fi un celular conectado por Wi-Fi.

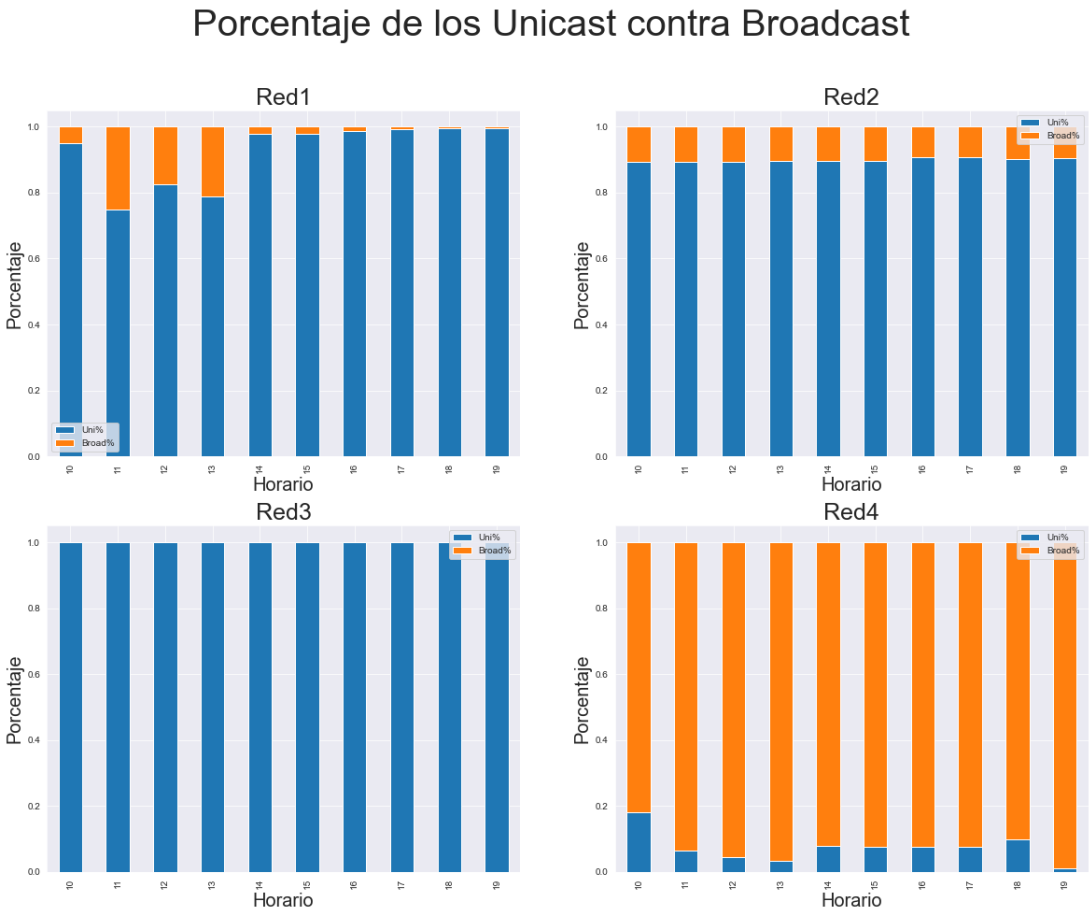
3. Experimentación

3.1. Unicast vs Broadcast

En esta subsección, tomaremos los datos recolectados de cada red y realizaremos comparativas según el tipo de transmisión de los paquetes capturados. De esta manera, por un lado tendremos todos los paquetes de tipo Unicast sin importar su número de protocolo, y por el otro tendremos todos los de tipo Broadcast, sin importar nuevamente qué protocolo vino en el símbolo.

A continuación presentamos 4 gráficos, uno por cada una de las redes descriptas en la sección anterior, para las cuales realizamos capturas de paquetes entre las 10 y las 19 horas de distintos días. En azul Unicast y en naranja Broadcast.

Antes de haber obtenido los resultados, nuestra intuición nos decía que las proporciones en todas las redes no iban a ser demasiado dispares ya que se trata de 4 redes domésticas bastante similares.



Como podemos observar, en las redes 1 y 2 tenemos aproximadamente el mismo porcentaje de paquetes de tipo Broadcast en promedio a lo largo de las horas. La única diferencia es que en la Red2, el porcentaje de paquetes de Broadcast está distribuido uniformemente a lo largo de las horas, mientras que en la Red1 hay una gran concentración de paquetes de este tipo entre las 11 y las 13 horas.

En contraposición, tenemos las redes 3 y 4 que tuvieron un comportamiento prácticamente opuesto entre ellas. Cabe destacar que la captura de datos de la Red3 tuvo lugar un domingo en el cual el uso de internet fue mínimo. Esto puede explicar por qué prácticamente no se realizaron envíos de tipo Broadcast. Los ARP request no fueron necesarios en este horario ya que las direcciones MAC indispensables para un acceso mínimo a internet ya se encontraban en la caché correspondiente al protocolo ARP, y prácticamente no hubo interacción entre dispositivos. Por otro lado, en la Red4 encontramos una relación completamente distinta a las demás redes. Este último análisis contradice nuestra hipótesis en la que preveíamos unos resultados mucho más similares. A priori no encontramos una explicación para justificar estas diferencias tan notorias. Tal vez sería interesante hacer un análisis en profundidad de esa red contrastada con cualquiera de las otras, teniendo en cuenta las sutilezas de herramientas de asignación de IP como DHCP.

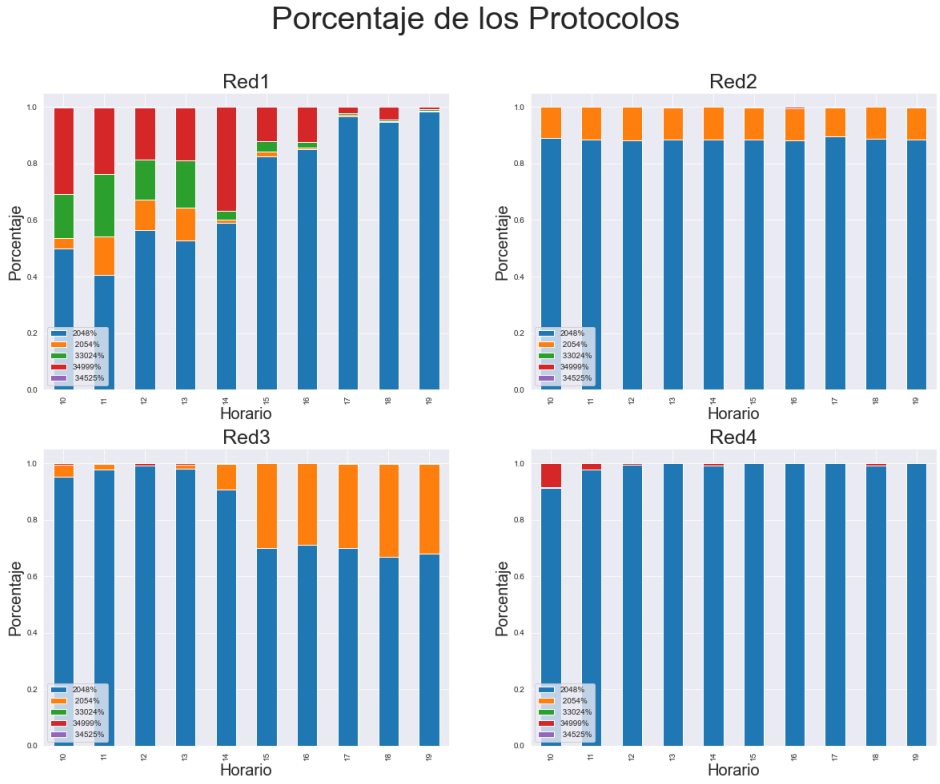
3.2. Protocolos encontrados

Para este experimento vamos a analizar qué protocolos se encontraron en las distintas redes y ver cuáles son los protocolos que predominan en éstas. Es por esto que primero que vamos a identificar cada protocolo y su funcionamiento a continuación.

- **IPv4 (2048):** IPv4 funciona en la capa de red de la pila de protocolos TCP o IP. Su tarea principal es principalmente transferir los bloques de datos desde el host de envío al host de destino, donde los remitentes y los receptores son ordenadores que se identifican de forma única por las direcciones de protocolo de Internet.
- **ARP (2054):** Es un protocolo de comunicaciones de la capa de enlace de datos, responsable de encontrar la dirección de hardware (Ethernet MAC) que corresponde a una determinada dirección IP.
- **Customer VLAN Tag Type (33024)**: Este protocolo está pensado para darle soporte a una LAN virtual en una red Eethernet 802.3.
- **IPv6 (34525):** IPv6 es una actualización al protocolo IPv4, diseñado para resolver el problema de agotamiento de direcciones.
- **IEEE Std 802 - OUI Extended Ethertype (34999):** El protocolo IEEE 802.1Q fue un proyecto del grupo de trabajo 802 de IEEE para desarrollar un mecanismo que permita a múltiples redes interconectadas con puentes

o switches compartir transparentemente el mismo medio físico sin problemas de interferencia entre las redes que comparten el medio. Permite identificar a una trama como proveniente de un equipo conectado a una red determinada. Una trama perteneciente a una VLAN sólo se va a distribuir a los equipos que pertenezcan a su misma VLAN, de forma que se separan dominios de broadcast.

Nuestra hipótesis para este experimento es que los protocolos que deberían predominar en la gráfica son los de IPs ya que estos protocolos transmiten datos de usuario mientras que los otros son protocolos de control en general.



Podemos ver que en los 4 gráficos, efectivamente el protocolo con más presencia es el protocolo IP. En las redes 3 y 4 hubo una minoritaria aparición de paquetes con protocolo ARP, y en la Red4 hubo una presencia no negligible de paquetes con protocolo 34999.

Cabe destacar la variedad de protocolos encontrados en la Red1. Esto se debió al uso intensivo de la red: a diferencia de las demás redes, la Red1 ha sido usada para una amplia cantidad de operaciones en línea tales como las relacionadas a youtube, spotify, descarga de archivos y hasta participación en una comunidad de Twitch.

3.3. Entropía de las redes

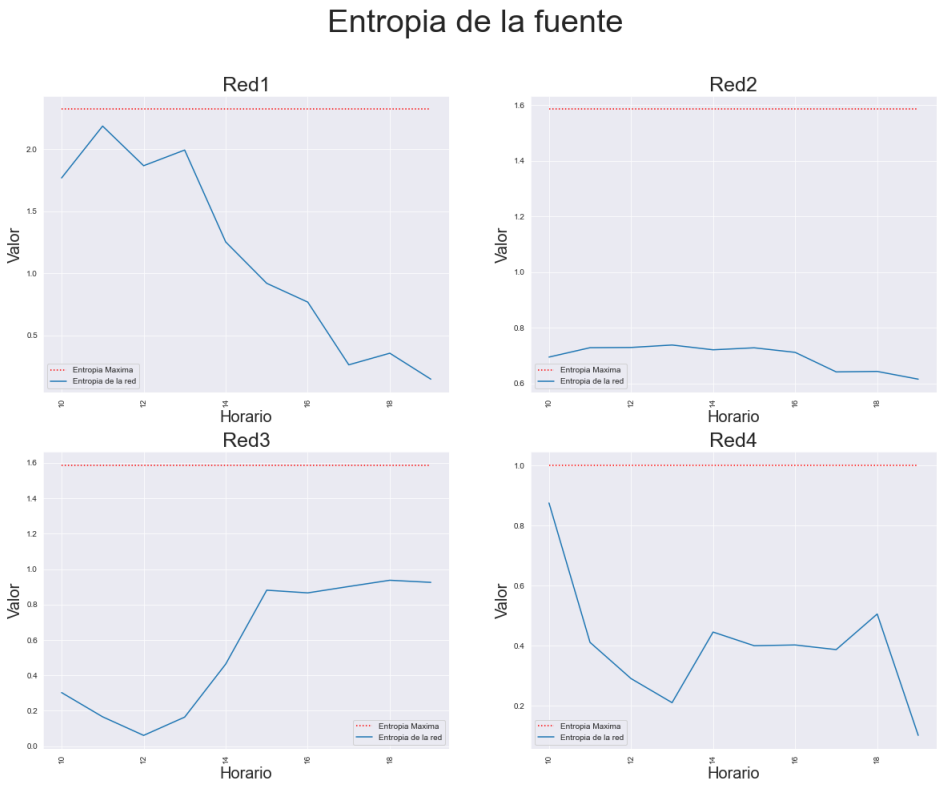
Esta subsección corresponde a los datos pertinentes a la entropía de cada una de las redes. Es por esto que presentaremos 4 gráficos mostrando la cantidad media de información por símbolo de cada una de las fuentes modeladas entre las 10 y las 19 horas. Además, en cada gráfico tendremos la entropía máxima de la red correspondiente para así ver si la entropía empírica se acerca a la máxima.

A modo de continuar con el análisis integral comenzado en las subsecciones anteriores, podemos hipotetizar que para las redes 2, 3 y 4 tendremos una entropía muy baja: en la sección 3.2, pudimos observar que en estas redes hubo muy poca variedad de protocolos ya que predominó el protocolo IP. La Red2 será la de menor entropía dada la aparición estable y minoritaria de paquetes con protocolo ARP, mientras que el resto de los paquetes son IP. Esto, combinado a que en la sección 3.1 la Red2 tiene mayoritariamente paquetes Unicast, resulta en una poca variedad de símbolos y por ende poca información brindada por estos. Esto provoca que la entropía de la Red2 sea la más baja de las 4.

Por lo visto en las dos subsecciones anteriores, las redes 3 y 4 tienen comportamientos similares a la Red2, aunque poseen un poco más de variedad ya sea en el número de protocolo o en el tipo de transmisión. Es por esto que tanto la Red3 como la Red4 presentarán una entropía moderada pero más alta que la obtenida para la fuente de la Red2.

El caso de la Red1 será completamente distinto: si bien en la sección 3.1 presenta poca variedad entre Unicast y Broadcast, en la sección 3.2 tenemos una gran variedad de números de protocolo, principalmente entre las 10 y las 14 horas. Esto lleva a una gran variedad de símbolos presentes con probabilidades más equitativas que las presentes en las redes anteriores y por ende una mayor entropía.

A continuación, presentamos los 4 gráficos correspondientes a las 4 redes analizadas:

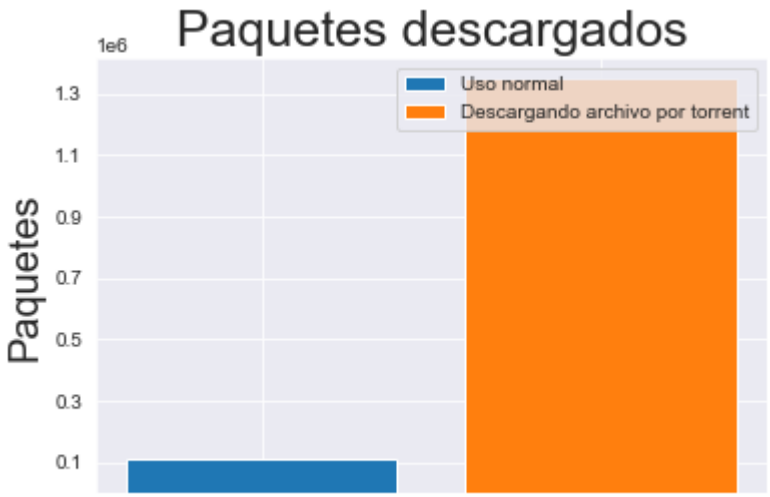


Como podemos ver, la Red1 alcanza una entropía muy alta entre las 10 y las 14 horas, llegando a acercarse considerablemente a la entropía máxima a las 11 horas. Sin embargo, con el paso de las horas, esta entropía disminuyó. Esto se condice con la baja en la variedad de protocolos (es decir, en la variedad de símbolos) que hubo en esta red a partir de las 14 horas.

Por otro lado, la Red2 ha sido la de menor entropía y las redes 3 y 4 han presentado una entropía más moderada, menor que la Red1 pero mayor que la Red. Esto confirma nuestra hipótesis.

3.4. Descarga de archivos

Para este experimento utilizamos la Red1 y corrimos el script de sniff nuevamente pero esta vez descargando una película de 2.7GB via torrent, nuestra hipótesis es que la cantidad de paquetes que se obtienen es mucho mayor cuando se descarga algún archivo que cuando se navega sin más.



Como muestra el gráfico, la cantidad de paquetes descargados es mucho mayor cuando se baja archivos. Es importante aclarar que la comparación esta dada en la misma hora del día e igualmente este supera por mucho la cantidad máxima descargada de paquetes del día de la medición original.

3.5. Modelado de la fuente S2

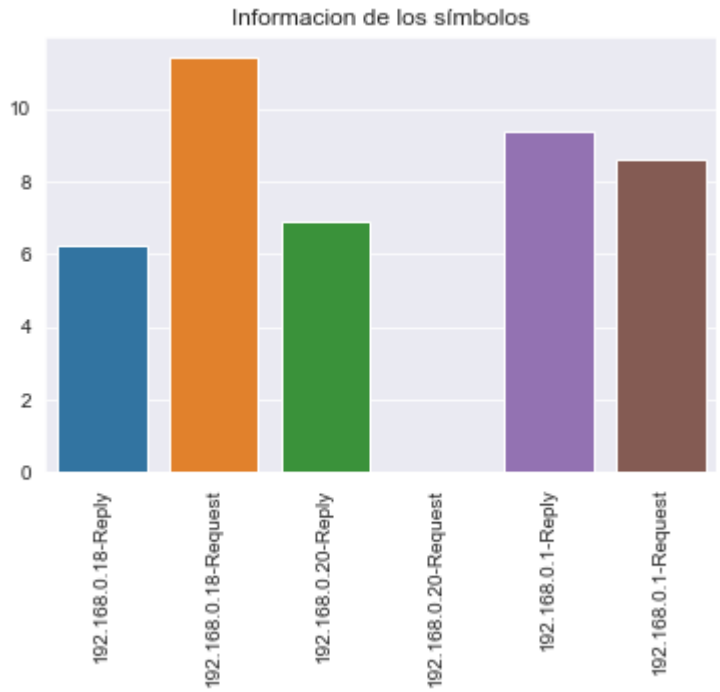
Para este último experimento, decidimos que cada símbolo sea de la forma <IP emisor, IP receptor, REQUEST | REPLY>. Esta elección se debe a que, a diferencia de la fuente S1, ya no nos importa destacar el número de protocolo puesto que siempre será 2054. Además, nos importa ahora la dirección del emisor y no solo la del receptor. Por otra parte, en lugar de hablar del tipo de transmisión (Unicast o Broadcast), consideramos más declarativo mencionar en qué parte del proceso del protocolo ARP se encuentra el paquete capturado (Reply o Request).

De esta forma, podremos determinar que un nodo distinguido será el que más información haya proveido. Como las 4 redes a las que tenemos acceso son domésticas, nuestra hipótesis es que el nodo distinguido siempre será el router ya que la principal acción ejecutada por los Hosts es conectarse a internet: para esto, la primera vez, deben pasar por el router y entonces estos Hosts harán ARP Request en busca de la MAC del router mediante un Broadcast. Pero luego de esto, salvo excepciones, siempre tendrán la MAC del Router en su caché de ARP, por lo que no será necesario efectuar otro Request en busca de la MAC del router y entonces el router no tendrá que realizar otro Reply. Esto llevaría a que la información provista por el router sea significativamente mayor que la de los demás nodos.

Presentaremos el gráfico de una sola red dado que en las otras 3 había muy

poco tráfico (o incluso nulo) de paquetes con protocolo ARP, por lo que no era posible realizar un análisis significativo. A continuación, una figura comparativa en la cual las direcciones IP utilizadas corresponden a:

- 192.168.0.18: la computadora desde la cual se corrió el script.
- 192.168.0.1: el router.
- 192.168.0.20: el extensor de señal.



Aquí podemos ver que, si tenemos en cuenta los replies y los requests hechos por estos dispositivos, la computadora es el nodo que más información aportó. Esto se debe a que realizó muy pocos replies y aún menos requests. Esto terminó brindando más información que el resto. Sin embargo, teniendo en cuenta solo los paquetes de tipo Reply, el que más aporta es el router. Esto se debe, ahora sí, a que no necesita seguir realizando replies porque los demás dispositivos cuentan con una caché para MACs, y allí pueden guardar la correspondiente al router.

4. Conclusiones

A lo largo de este informe, hemos analizado integralmente 4 redes domésticas. Como hemos visto en las distintas secciones, la naturaleza de estas redes no nos han permitido obtener una gran variedad de símbolos en general. Sin

embargo, consideramos que con lo que obtuvimos pudimos comprender el comportamiento general de una red doméstica a nivel de enlace.

Por otro lado, dado que las características de las 4 redes son levemente distintas, no hemos encontrado correlación entre estas características y las entropías calculadas. De hecho, vimos la sección de experimentación que la entropía se relaciona más con el uso de la red que con sus características. Esto se veía reflejado en el uso intensivo y variado de la Red1 entre las 10 y las 14 horas, lo que traía más variedad de símbolos y más información, es decir, más entropía. Esta red fue la que presentó la entropía más cercana a su entropía máxima teórica, aunque no ha logrado alcanzarla. El resto de las redes no logró acercarse salvo la Red4 al principio de todo, aunque luego su entropía disminuyó abruptamente.

En el análisis llevado a cabo, pudimos observar los protocolos más comunes transmitidos por nuestras redes. En general, las 4 redes presentaron los mismos tipos de protocolo, salvo la Red1. En esta, aparecieron 3 tipos de protocolos que no se hallaron en las demás redes: IPv6 (34525) y Customer VLAN Tag Type (33024). Ambos ya han sido descriptos en la sección 3.2 correspondiente a la experimentación con los protocolos encontrados.

Como comentario final nos pareció extremadamente interesante la diferencia encontrada entre la red 4 y las demás en cuanto al tipo de transmisión. Daría para un análisis más exhaustivo de esa red para ver si se pueden replicar los resultados y averiguar el motivo de esa diferencia abismal en las proporciones.