



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Sanguijuelas

3 de septiembre de 2014

Métodos Numéricos
Trabajo Práctico Nro. 1

Integrante	LU	Correo electrónico
Martin Carreiro	45/10	martin301290@gmail.com
Kevin Kujawski	459/10	kevinkuja@gmail.com
Juan Manuel Ortíz de Zárate	403/10	jmanuoz@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. Resumen	3
2. Introducción teórica	4
2.1. Algoritmo de eliminación gaussiana	4
2.2. Matriz banda	5
3. Desarrollo	6
4. Resultados	7
5. Discusión	8
5.1. Red del Alto Palermo	8
5.2. Red de Subway	8
5.3. Red de Casa	8
5.4. Red de McDonald's	8
6. Conclusiones	10
7. Referencias	11

1. Resumen

Mediante el manejo de matrices se buscará modelar y solucionar un problema de la realidad. Cómo esta esta compuesta de infinitas variables dicha modelización implicará una inevitable discretización. Es decir trabajar con una cantidad acotada de variables del problema (sólo las relevantes). Si bien el problema en cuestión consiste en decidir cual/es de las sanguijuelas que están adheridas al parabrisa se debe eliminar la modelización del mismo no es trivial. Es más, podría decirse que este proceso es mucho mas complejo y costoso que la solución en sí. ¿Por qué? porque la creación de la matriz que represente al parabrisa y el cálculo de las temperaturas en cada punto, si no se usa buen método, podría llegar a demorar mucho tiempo, tirar overflows o directamente nunca terminar.

Por esto es que a lo largo de este tp haremos mucho foco en como calcular las temperaturas, como optimizar el espacio ocupado por la matriz obtenida y como optimizar lo mas posible todas las operaciones matriciales.

2. Introducción teórica

Dado que la modelización utilizada para este problema es una matriz y que en base a algunos datos que nos dan de ella (posición de las saniguijuelas, dimensiones y temperaturas de los bordes) tenemos que calcular los valores de muchas celdas, lo que hay que resolver es un sistema de ecuaciones. Lo primero que se nos viene a la mente cuando tenemos esto es el método de eliminación gaussiano. Este transforma la matriz de coeficientes en una matriz triangular superior y luego mediante back substitution se pueden obtener todos sus valores.

Veamos ahora más en detalle como funciona cada uno de estos algoritmos.

2.1. Algoritmo de eliminación gaussiana

Es importante en este punto aclarar que para poder hacer estos cálculos la matriz debe ser cuadrada (7). Los pasos a seguir en este algoritmo son:

- Ir a la columna no cero extrema izquierda
- Si el primer renglón tiene un cero en esta columna, intercambiarlo con otro que no lo tenga
- Luego, obtener ceros debajo de este elemento delantero, sumando múltiplos adecuados del renglón superior a los renglones debajo de él
- Cubrir el renglón superior y repetir el proceso anterior con la submatriz restante. Repetir con el resto de los renglones (en este punto la matriz se encuentra en la forma de escalón)
- Comenzando con el último renglón no cero, avanzar hacia arriba: para cada renglón obtener un 1 delantero e introducir ceros arriba de éste sumando múltiplos correspondientes a los renglones correspondientes

Ejemplo:

$$\begin{cases} 2x & +y & -z & = & 8 \\ -3x & -y & +2z & = & -11 \\ -2x & +y & +2z & = & -3 \end{cases}$$

Figura 1: Sistema de ecuaciones de 3x3

$$\begin{cases} 2x & +y & -z & = & 8 \\ & \frac{1}{2}y & +\frac{1}{2}z & = & 1 \\ & 2y & +z & = & 5 \end{cases}$$

Figura 2: Sistema de ecuaciones con ceros en la primera columna

$$\begin{cases} 2x & -2z & = & 6 \\ & \frac{1}{2}y & + \frac{1}{2}z & = & 1 \\ & & -z & = & 1 \end{cases}$$

Figura 3: Sistema de ecuaciones triangulado

En esta última imagen podemos observar que el valor de Z ya lo tenemos. Por lo tanto si reemplazamos el mismo en la ecuación arriba obtenemos Y y lo mismo con la primer ecuación.

La complejidad computacional de esto es aproximadamente n^3 . Esto es, el número de operaciones requeridas es n^3 si el tamaño de la matriz es $n \times n$. Por lo tanto puede volverse un cálculo sumamente grande si la matriz es de dimensiones importantes. Por esto decidimos pensar un poco mejor la situación y encontrarle la vuelta para resolverlo en un tiempo razonable.

2.2. Matriz banda

Analizando cuidadosamente la ecuación del cálculo de cada celda EXPLICACION BANDA

3. Desarrollo

Como se mencionó anteriormente, el objetivo es analizar la estructura de algunas redes locales mediante la captura de paquetes ARP que se envían por esas redes. Estos paquetes servirán para obtener información sobre los dispositivos de red y sacar conclusiones de los resultados que se consigan.

Lo primero que se hace en el trabajo es implementar una herramienta para escuchar de manera pasiva una red local. La idea es que la red local que analicemos no sea una red controlada, para favorecer un análisis más rico (si fuera controlada, sabríamos por ejemplo cuál es la IP del router). Para eso se utiliza *Scapy* una biblioteca escrita en *Python*, para la captura y manipulación de paquetes en redes.

El segundo paso es definir los siguientes modelos de fuente de información:

- $S_{src} = \{s_1, \dots, s_k\}$ donde cada símbolo s_i es una dirección IP que aparece como dirección origen en los paquetes ARP de tipo *who-has*.
- $S_{dst} = \{d_1, \dots, d_k\}$ donde cada símbolo d_i es una dirección IP que aparece como dirección destino en los paquetes ARP de tipo *who-has*.

Con estas fuentes definidas, ya se puede estimar las probabilidades de cada IP que necesitemos de los paquetes ARP capturados y calcular la cantidad de información de cada uno para luego conocer la entropía de cada fuente. Para estimar las probabilidades de las IPs en las fuentes, lo que se hace es simple. Supongamos que *#paquetes* es la cantidad total de paquetes ARP *who-has* obtenidos y *#apariciones* es la cantidad de veces que aparece una *ip* determinada (la explicación sirve para ambas fuentes de información), entonces se define la probabilidad de esa *ip* como:

$$P(ip) = \frac{\#apariciones}{\#paquetes}$$

Con esta definición se puede calcular la cantidad de información de cada IP obtenida en las fuentes. Luego, es fácil obtener la entropía de cada fuente de información.

Entonces, haciendo uso de esta herramienta implementada, se deben realizar capturas de paquetes ARP sobre distintas LANs (*Local Area Network*) para poder hallar nodos (dispositivos de red) distinguidos. El router que oficia como *gateway* (se comunica con otras redes) de la LAN analizada es de particular interés. La probabilidad de que su IP aparezca en un paquete ARP *who-has*, sea como emisor o como destinatario debería ser alta. Es decir, en teoría debería ser la IP más frecuente en este tipo de paquetes porque es el canal común de comunicación. Otra manera de verlo, es la siguiente: se busca la dirección IP (notar que puede haber más de un *gateway*) cuya probabilidad sea la más alta y por tanto, es la que menor cantidad de información aporta. Se supone que las direcciones IP cuya información sea más cercana al valor de la entropía de la fuente serán los *gateways* de la red.

Asimismo, además de analizar la LAN para poder identificar el o los *gateways*, se monitorea la actividad de las otras IPs que se referencian en la red para registrar la frecuencia de cada una, comparando su información con la entropía de la fuente y quizás, encontrar otros nodos especiales.

4. Resultados

5. Discusión

Espiral vs centrico

no contar los bordes ni las sanguijuelas para crear la matriz banda

5.1. Red del Alto Palermo

En esta red se observa una gran interacción con el router, cuya IP es 172.17.0.1. Esto sucede porque el router es el dispositivo por el que pasan en general la mayoría de los pedidos. Un aspecto interesante que se ve en esta comunicación es que no parece haber un pedido fuera de la red, es decir, a otro dominio. Lo que se puede deducir es que el dominio de esta red parece ser 172.17.0.0 /16 donde la IP del router es 172.17.0.1, el cuál se comunica con el resto de los dispositivos bajo este dominio.

5.2. Red de Subway

Lo primero que se ve es la baja cantidad de IPs que interactúan en la comunicación. Además se observa que la interacción se realiza en una red privada del dominio 192.168.0.0 /16 que se suele usar para redes domésticas. Naturalmente la IP del router en este tipo de red es 192.168.0.1.

De los datos obtenidos, se observó que casi todos los pedidos tenían como destinatario la IP del router. Algo que se puede ver en el histograma realizado para esta red es que la IP que menos información tiene es 192.168.1.20, lo que significa que es la que más pedidos ARP efectuó.

5.3. Red de Casa

La primera observación que se puede realizar es que a pesar de contar únicamente con 5 diferentes dispositivos con acceso a la red (incluyendo el router) se pueden observar un total de 13 diferentes IPs. En segunda instancia se puede detectar fácilmente la línea de la entropía en un valor aproximado de 2,1 y así también la única IP que aporta menor información que la entropía a la IP del router en valor de 192.168.0.1 el cual da salida hacia internet para los dispositivos de la red. Notar que el dominio de la IP del router vuelve hacer la red doméstica 192.168.0.0 /16, lo que es natural porque el experimento fue realizado en un departamento.

5.4. Red de McDonald's

Esta red es la que más datos aportó al análisis, fue claramente en la que más pedidos ARP se hicieron entre distintos nodos. Esto hace que tenga algunos aspectos interesantes para observar, en comparación con las otras redes analizadas. A diferencia de las anteriores redes se puede destacar que al ser una red de mayor incidencia (contar con una cantidad alta de dispositivos en la comunicación), la diferencia de información entre la IP de menor información (el router) y la entropía de la fuente es mayor.

Lo que no se observa bien en el histograma es que la mayoría de las IPs son del estilo 172.17.203.X, lo que nos induce a pensar que el dominio de la red es 172.17.203.0 \24 y que la mayor interacción se produce en ese dominio, en donde la mayoría de los pedidos ARP tienen como destinatario la IP 172.17.203.1 que se deduce que la IP correspondiente al router de la LAN (es la que menos información aporta en la fuente graficada en el histograma, es decir, es la barra más pequeña).

Otro aspecto interesante que se observó es la gran cantidad de veces que figuraba 0.0.0.0 como IP emisora. Este es un pedido ARP gratuito que puede ser utilizado por cualquier dispositivo para verificar que determinada IP no esté siendo usada (la IP destino del pedido). La aparición de esta IP tiene relación con el siguiente hecho: la IP destino de estos pedidos siempre pertenecían al dominio 169.254.0.0 /16, usada como *broadcast* de *link local*. Se usa para comunicación entre hosts en un sólo *link*. Los hosts obtienen direcciones de este dominio mediante una autoconfiguración, por ejemplo, en situaciones donde el servidor DHCP (*Dynamic Host Configuration Protocol*) no se encuentra. Es decir, cuando una interfaz pierde la conexión o es activada, si no puede recibir una IP con el protocolo DHCP se le asigna una IP del dominio 169.254.0.0 /16. Por lo tanto, con el pedido de parte de 0.0.0.0 se quiere fijar si la IP que le fue asignada no está siendo usada en la red. El último comentario con este dominio, que también está relacionado con el pedido gratuito, es el siguiente: se observó un pedido ARP donde la IP emisora era la misma que la destinataria por cada IP del dominio 169.254.0.0 /16. Esta particularidad la vimos en el taller; cuando se levanta una interfaz muchos sistemas envían automáticamente un pedido ARP gratuito con esta característica. De esta manera se pueden detectar IPs duplicadas (en caso de recibir respuesta) y actualiza la caché ARP de los demás hosts de la red.

6. Conclusiones

Se pudo observar que capturar los paquetes ARP en distintas redes locales da bastante información sobre la red. Es decir, con un análisis estadístico sobre los paquetes ARP *who-has* tomando como referencia las fuentes de información que se pueden definir a partir de las IPs que emiten mensajes y las IPs destinatarias, se puede por ejemplo deducir cuál es el router de la red. Otra manera de verlo es que si ya se sabe cuál es la IP del router, los experimentos sirven para ver que efectivamente el router es el dispositivo más solicitado, como se puede ver en la experimentación que se realizó en el departamento de un miembro del equipo.

Se ejecutó la herramienta implementada en distintas redes, lo que supuso distintos resultados. Lo interesante de esto es detectar nodos distinguidos o un comportamiento especial que hizo que tuviéramos que averiguar ciertas cosas, como sucedió en las capturas de la red del McDonald's en donde aparece una forma de asignar IPs cuando se levantan interfaces. Asimismo, con los histogramas hechos de las fuentes correspondientes se puede ver el grado de actividad de ciertos dispositivos en la red (por lo menos en el momento en que fue realizado el experimento; para mayor precisión se debería realizar una captura más amplia y de más duración).

Como última observación cabe destacar que no se debió participar 'agresivamente' en la red. Es decir, la escucha de los paquetes ARP fue pasiva. Sólo con recopilar información durante un tiempo determinado se puede deducir con bastante confiabilidad qué dispositivo hace de servidor para el resto de los nodos de la red.

7. Referencias

- Una matriz de n por m elementos, es una matriz cuadrada si el número de filas es igual al número de columnas, es decir, $n = m$ y se dice, entonces que la matriz es de orden n
- En matemáticas una matriz se le llama matriz banda cuando es una matriz donde los valores no nulos son confinados en un entorno de la diagonal principal, formando una banda de valores no nulos que completan la diagonal principal de la matriz y más diagonales en cada uno de sus costados.