



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico

Rutas en internet

Teoría de las Comunicaciones

Integrante	LU	Correo electrónico
Verma, Tusar	309/22	tusarverma12345@gmail.com

Instancia	Docente	Nota
Primera entrega		
Segunda entrega		



Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón Cero + Infinito)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Conmutador: (+54 11) 5285-9721 / 5285-7400

<https://dc.uba.ar>

1. Introducción

En este trabajo vamos a desarrollar una herramienta similar a traceroute y la usaremos para conectarnos a distintas universidades en varias partes del mundo con el objetivo de detectar los saltos interoceánicos. Para ello analizaremos los RTTs que se obtienen en cada hop.

Luego extenderemos la herramienta con la propuesta de Cimbala para estimación de outliers, para poder inferir cuales hops corresponden a dichos saltos y lo compararemos con el análisis manual de RTTs.

2. Métodos y condiciones de los experimentos

Se implementaron 3 funciones para el calculo de todos los datos del experimento.

- *calcular_rtts_time_exceeded()*

Esta función toma una IP y manda paquetes ICMP hasta el destino incrementando el TTL del paquete iterativamente. Ésto con el objetivo de poder trazar la ruta que toman los paquetes. En general, no todos los nodos de la ruta son visibles. Algunos motivos podrían ser que los nodos no implementan ICMP o se tenga algún firewall que bloquee los paquetes.

Este proceso se repite varias veces para obtener así nuestra muestra de RTTs por TTLs.

La función devuelve por cada ttl donde se recibió al menos una respuesta, una lista de (IP, RTT) donde IP corresponde al que respondió el mensaje de TTL time exceeded y RTT es el tiempo transcurrido entre el envío del paquete y la vuelta del mensaje de error.

- *calcular_rtt_promedio_por_ttl()*

Por cada TTL puede haber distintas nodos que hayan respondido por las distintas rutas que los paquetes podrían haber tomado. Por eso esta función toma aquel que mayor respuestas haya tenido y promedia sus RTTs.

Es decir, la función devuelve una lista, donde por cada TTL donde hubo una respuesta se tiene el par (IP, RTT promedio), donde IP es el nodo que mayor respondió para dicho TTL.

- *calcular_rtt_entre_saltos()*

Finalmente se calcula la diferencia entre RTTs en saltos que deberían ser consecutivos por el TTL.

Para los experimentos se seleccionaron las siguientes 5 universidades:

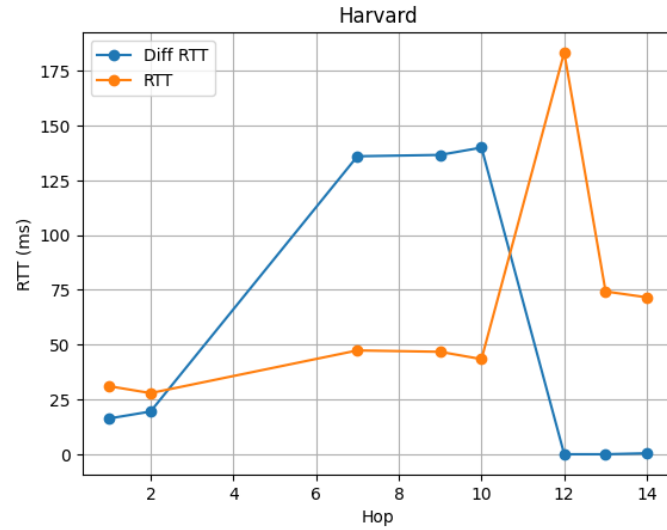
- Universidad Harvard en Estados Unidos, San Francisco: 192.0.66.20
- Escuela Politécnica Federal de Zúrich en Suiza, Zurich: 129.132.19.216
- Universidad Técnica de Múnich en Alemania, Munich: 129.187.255.109
- Universidad de Nairobi en Kenya: 41.204.161.206
- Universidad de Copenhagen en Dinamarca, Copenhagen: 130.226.237.173

3. Resultados

En los siguientes gráficos podemos observar los resultados de los experimentos. En naranja tenemos los distintos valores de RTT para los TTLs donde hubo respuesta. Es decir, para los 25 saltos que se esperaba calcular, en ninguno de los casos obtuvimos respuestas en todos los hops hasta el destino.

En azul tenemos la diferencia en RTT entre el salto siguiente y el actual. Esta diferencia podría ser negativa si el hop siguiente tuvo un RTT menor, dependiendo de la ruta que el paquete tomó.

3.1. Harvard

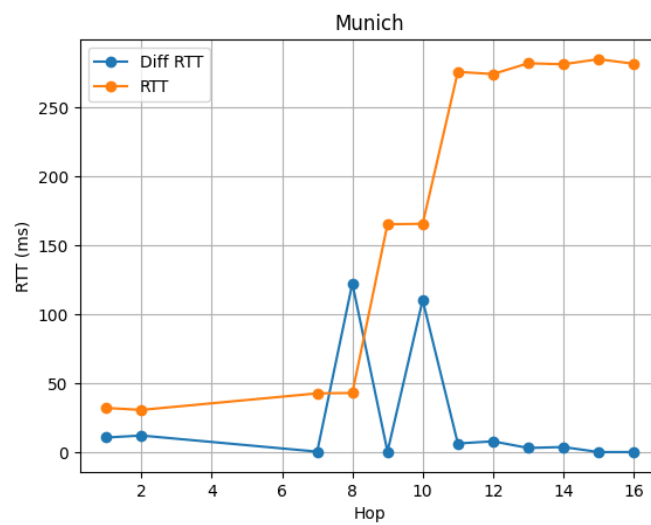


Para el experimento con la comunicación con Harvard se obtuvo los siguientes datos:

TTL	IP	Pais	RTT
1	10.0.2.2	Local	31.03ms
2	192.168.0.1	Local	27.87ms
7	181.96.113.234	Argentina	47.35ms
9	181.96.113.217	Argentina	46.73ms
10	181.96.113.217	Argentina	43.39ms
12	200.25.51.246	Colombia	183.31ms
13	200.25.57.147	Brasil	74.27ms
14	192.0.66.20	EEUU	71.5ms

Como se puede observar en la tabla, se obtuvo un RTT menor para el TTL 14 que llega hasta el destino. Y hubo un salto relativamente grande entre TTL 10 y TTL 12. Esto podría haberse causado por alguna congestión en la zona, que luego en siguientes iteraciones y con más saltos, se eligió otro camino con menos retardo.

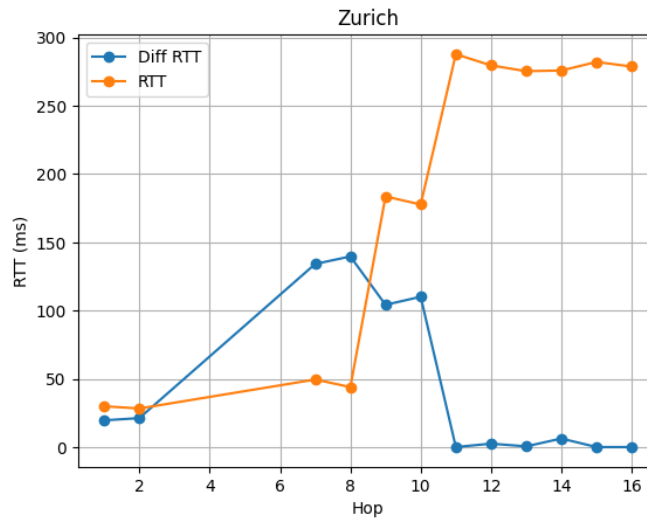
3.2. Munich



TTL	IP	Pais	RTT
1	10.0.2.2	Local	32.04ms
2	192.168.0.1	Local	30.58ms
7	181.89.51.39	Argentina	42.58ms
8	195.22.220.56	Argentina	42.89ms
9	89.221.41.221	EEUU	165.31ms
10	4.68.62.57	EEUU	165.62ms
11	171.75.10.131	Alemania	275.88ms
12	62.67.67.154	Alemania	274.32ms
13	188.1.145.230	Alemania	282.10ms
14	188.1.37.90	Alemania	281.42ms
15	129.187.0.168	Alemania	285.08ms
16	129.187.255.109	Alemania	281.77

En este caso podemos notar una clara correlación entre diferencias grandes en el RTT de un TTL a otro (por ejemplo del 8 al 9 o del 10 al 11) y saltos interoceánicos.

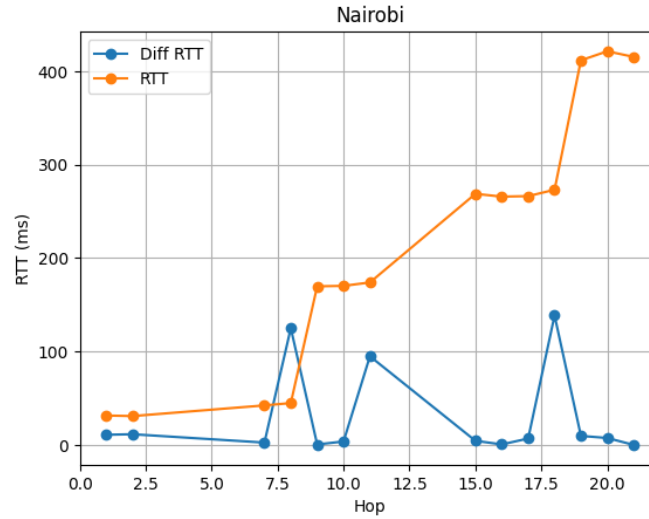
3.3. Zurich



TTL	IP	Pais	RTT
1	10.0.2.2	Local	29.89ms
2	192.168.0.1	Local	28.26ms
7	181.96.113.234	Argentina	49.47ms
8	195.22.220.56	Argentina	43.97ms
9	89.221.41.1	EEUU	183.64ms
10	89.221.41.193	EEUU	177.74ms
11	141.136.108.137	Alemania	287.86ms
12	154.14.130.98	Suiza	279.67ms
13	130.59.38.81	Suiza	275.41ms
14	192.33.92.1	Suiza	275.92ms
15	192.33.92.169	Suiza	282.17ms
16	129.132.19.216	Suiza	278.82ms

Podemos observar en este caso el mismo fenómeno que ocurrió en el caso anterior. Cuando saltamos de Argentina a EEUU, el RTT tuvo un incremento relativo mayor. Así como cuando pasamos de EEUU a los países europeos.

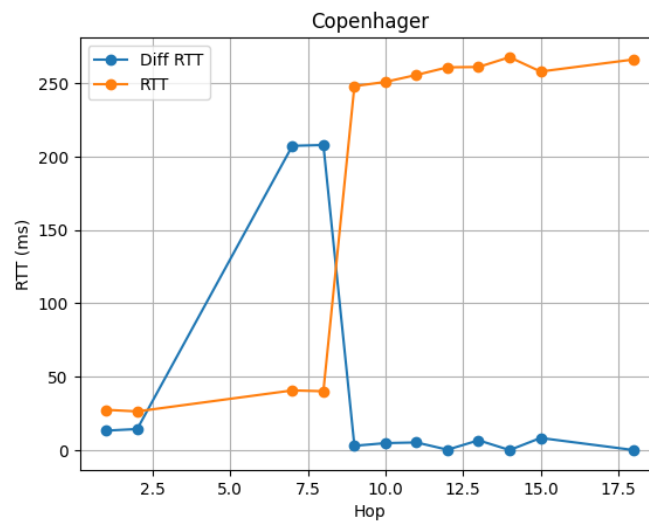
3.4. Nairobi



TTL	IP	País	RTT
1	10.0.2.2	Local	31.4264
2	192.168.0.1	Local	30.8460
7	181.96.113.234	Argentina	42.2787
8	195.22.220.56	Argentina	44.7662
9	89.221.41.221	EEUU	169.7054
10	195.22.199.33	EEUU	170.1950
11	129.250.2.109	EEUU	173.8633
15	129.250.3.249	UK	268.6987
16	129.250.4.141	UK	265.6953
17	83.231.146.66	UK	266.2332
18	196.32.210.57	Malawi	273.0659
19	197.136.176.58	Kenya	411.5910
20	197.136.168.17	Kenya	421.2122
21	41.204.161.206	Kenya	415.2505

En este caso tenemos 3 saltos grandes en el RTT: cuando pasamos de Argentina a EEUU, de EEUU a UK y de UK a Malawi. Es decir, pasamos por 3 continentes: América, Europa y África. Y en cada uno vemos un salto en el RTT (además del que va de Argentina a EEUU).

3.5. Copenhagen



TTL	IP	País	RTT
1	10.0.2.2	Local	27.4258
2	192.168.0.1	Local	26.3536
7	181.96.113.234	Argentina	40.6811
8	195.22.220.56	Argentina	40.1158
9	195.22.209.220	UK	248.1092
10	195.66.225.24	UK	251.0135
11	109.105.97.124	Holanda	255.7511
12	109.105.97.81	Dinamarca	260.9477
13	109.105.102.160	Dinamarca	261.1789
14	109.105.102.161	Dinamarca	267.8457
15	192.38.111.154	Dinamarca	258.0567
18	130.226.237.173	Dinamarca	266.3030

Siguiendo la misma tendencia que los 3 experimentos anteriores, pero ahora sin pasar por EEUU, cuando saltamos de un continente a otro el RTT se incrementa bastante.

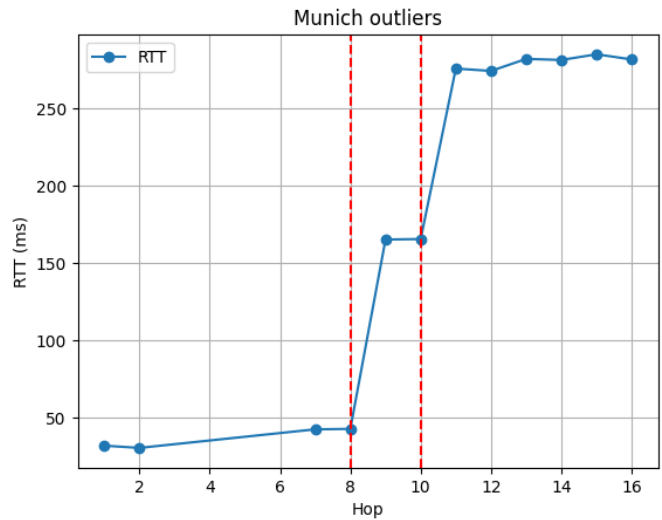
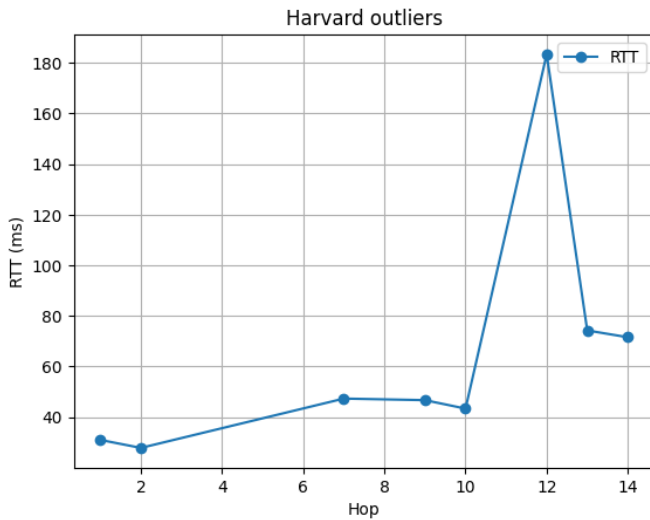
4. Outliers para detección de enlaces interoceánicos

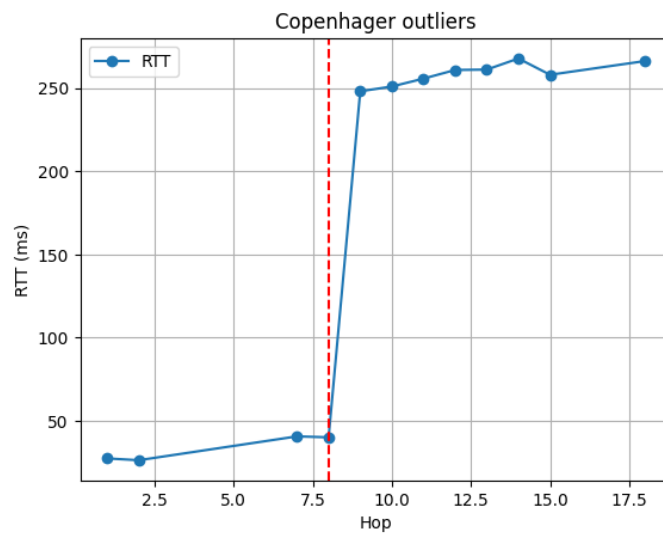
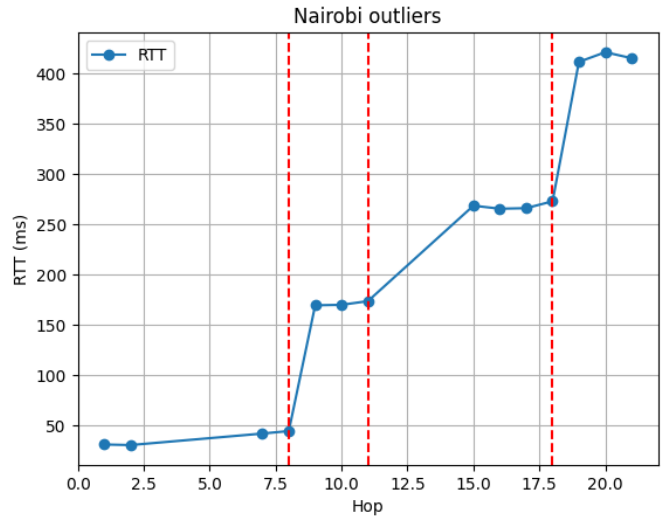
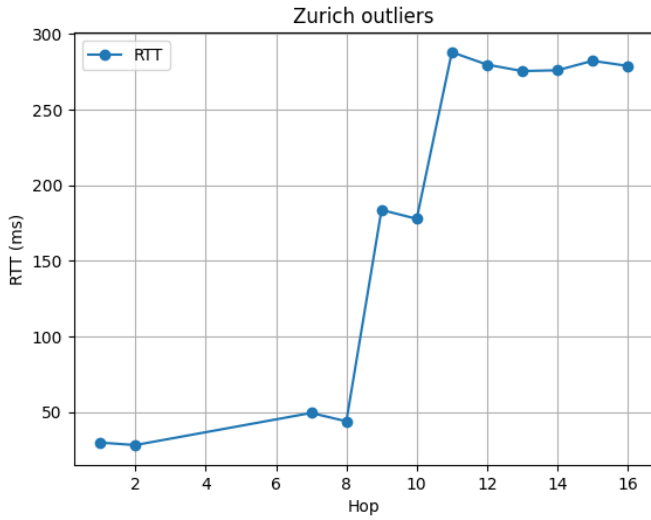
En esta sección utilizaremos el método planteado por Cimbalá para la detección de outliers con el propósito de poder predecir, usando las diferencias en el RTTs entre TTLs consecutivos, los saltos interoceánicos.

La idea es calcular la diferencia de RTTs entre un TTL y el siguiente. Con dicho conjunto de datos, buscamos aquellos que se alejen de la media. En general, es de esperarse que si un paquete viaja por un mismo país o dentro de un mismo continente, la diferencia de RTTs entre saltos no sea muy grande. Pero si se pasa a otro continente o a un país muy alejando, se esperaría que la diferencia de RTTs sea bastante grande. Éstos outliers en nuestro conjunto de datos son los que interpretaríamos como saltos interoceánicos.

En la siguiente experimentación, con los datos obtenidos en los experimentos anteriores, se calcula el conjunto de datos de diferencia de RTTs para cada universidad y se busca los outliers con el método de Cimbalá.

En las siguiente figuras se ve el resultado del experimento. Se muestra en azul los RTTs para los distintos TTLs obtenidos y en rojo los outliers que da el método de Cimbalá

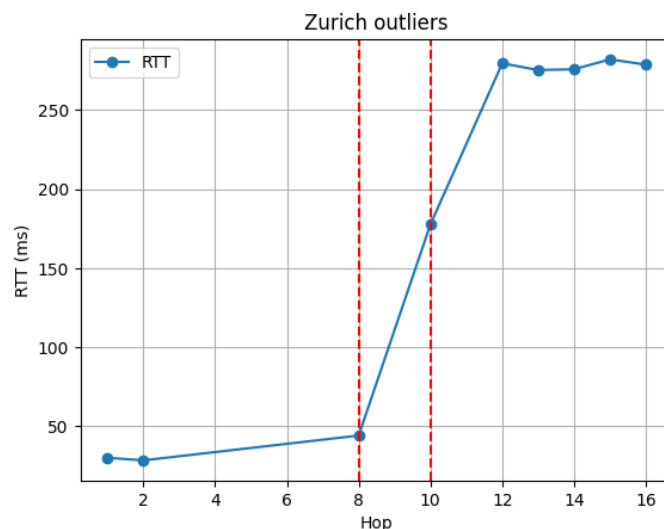




Podemos notar que para las universidades de Munich, Nairobi y Copenhagen efectivamente pudo detectar los saltos interoceánicos.

Para el caso particular de Harvard vemos que no pudo detectar el salto que se produce en el TTL 12. Esto podría deberse a que como en los datos hay muchos RTTs decrecientes, luego las diferencias se calculan con el próximo tal que sea mayor. Dando como resultado que las diferencias de los TTLs 7, 9 y 10 se calculen con el RTT del TTL 12. Luego la distribución de diferencias de RTTs no puede notar dicho salto.

Para el caso de Zurich vemos que tampoco pudo detectar los saltos, y podría deberse a una razón similar al de Harvard. Si eliminamos del conjunto algunos hops tal que no haya tantos datos decrecientes y los outliers no se oculten obtenemos:



Del experimento original se mantuvo los siguientes TTLs: 1, 2, 8, 10, 12, 13, 14, 15, 16. Y en este caso se pudo detectar los saltos correspondientes a Argentina-EEUU y EEUU-Alemania.

5. Conclusiones

En los distintos experimentos vimos como el cambio relativo en el RTT podría usarse para inferir saltos interoceánicos. Pero esto no siempre podría ser el caso. En particular, vimos en el experimento de la comunicación con la universidad de Harvard que tuvimos un salto grande en el RTT desde el TTL 10 al 12 pero éstos correspondían a nodos en el mismo continente (uno Argentina y otro Colombia). El motivo de este fenómeno podría deberse a alguna de las anomalías que el comando Traceroute sufre.

Como en el TTL 13 vemos como el RTT disminuye drásticamente y cambiamos a otro país (Brasil), muy probablemente se deba a que en la zona de Colombia por donde pasó el paquete haya estado congestionado. Luego en la siguiente iteración del algoritmo, el paquete viajó por alguna ruta menos congestionada.

En los otros 4 experimentos para las demás universidades pudimos obtener un resultado esperado, donde el RTT sufre un salto cuando pasamos a otro continente. Por lo que podría funcionar como estimador de saltos interoceánicos. Por eso se desarrolló, usando el método de Cimbala para calcular outliers, un estimador de saltos interoceánicos basado en las diferencias de RTTs.

Experimentando vimos que no siempre puede detectarlos, pero es un buen estimador. Los problemas provienen del lado de las anomalías del traceroute más que del método de detección de outliers.

Otro punto a destacar es que en ningún experimento tenemos la información completa de la ruta. Es decir, no tenemos información para todos los TTL. Ésto podría ocurrir porque dichos nodos no implementan ICMP, o hay algún firewall que no permite el envío de los paquetes o también alguna anomalía de traceroute.