DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 2

Rutas en Internet

22 de julio de 2016 — Teoría de las comunicaciones

Integrante	LU	Correo electrónico
Interlandi, Daniel	773/00	danielinterlandi@yahoo.com.ar
Ladelfa, Hernán Nahuel	318/04	nahueladelfa@gmail.com

Instancia	Docente	Nota
Primera entrega		
Segunda entrega		

Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria - (Pabellón I/Planta Baja)
Intendente Güiraldes 2610 - C1428EGA
Ciudad Autónoma de Buenos Aires - Rep. Argentina
Tel/Fax: (++54 +11) 4576-3300
http://www.exactas.uba.ar

Índice

1. Introducción

El objetivo de este trabajo es el de experimentar con las herramientas provistas por el protocolo ICMP para analizar las rutas tomadas por los paquetes hasta alcanzar su destino. Además se analizará la existencia de saltos entre nodos intercontinentales en las rutas de los experimentos.

2. Implementación

Para este trabajo implementamos nuestra propia versión de la herramienta `traceroute`, en lenguaje *Python* y utilizando la librería *Scapy*.
El funcionamiento general de nuestro programa es similar al del `traceroute` que se encuentra en los sistemas operativos más conocidos. Para lograr encontrar la ruta que podrían seguir los paquetes hasta alcanzar a un host destino lo que hacemos es ir enviando paquetes *ICMP Echo Request*, comenzando con un valor de TTL (Time To Live) igual a uno e incrementando este valor gradualmente. De esta forma, cada router al recibir estos paquetes decrementará el valor del TTL del paquete en uno. Si, luego de decrementar el valor del TTL, el mismo queda en cero, el router que posee el paquete responderá al host origen con un mensaje *ICMP Time Exceeded*. Cuando un paquete alcanza finalmente al host destino, este responderá con un mensaje *ICMP Echo Reply*. De esta forma, recibiendo los sucesivos mensajes *ICMP Time Exceeded* de cada router y con el mensaje final *ICMP Echo Reply*, podremos armar una ruta con los routers intermedios hasta el host destino.

Para evitar que el programa nunca finalice por intentar alcanzar al destino cuando este no responde, agregamos una cota al valor del TTL de 40. Además, para cada valor de TTL realizamos hasta 30 intentos para encontrar el correspondiente nodo y nos quedaremos con el promedio del valor de los RTT (Roundtrip Time). De esta forma podría darse la situación de que encontremos más de un nodo posible. En este caso tomaremos al que se presentó con mayor frecuencia y como RTT al promedio de los RTT obtenidos.

Además, a este programa le agregamos los cálculos necesarios para encontrar los enlaces entre nodos cuyos valores de RTT se encuentran considerablemente por encima del resto. Es decir, que buscaremos encontrar los outliers de las muestras. Estos cálculos los utilizaremos para inferir cuáles son los enlaces intercontinentales. Para encontrar a estos outliers utilizaremos la técnica de estimación de John M. Cimbala propuesta por la cátedra. Veremos más detalles sobre estos cálculos en la siguiente sección.

Para utilizar nuestra versión de `traceroute` se debe ejecutar:

```
$ python traceroute.py <host> <prefijo>
```

Donde:

- host: Es la dirección IP o nombre de dominio del host destino hasta el cual se quiere calcular la ruta.
- prefijo: Es el prefijo que se utilizará para generar los nombres de los archivos de salida.

3. Estimación de outliers

Basándonos en la técnica de estimación propuesta por John M. Cimbala para encontrar los outliers de una muestra, agregamos a nuestro programa los cálculos necesarios para inferir saltos intercontinentales en las rutas tomadas por los paquetes hacia un host destino.

Los pasos que seguiremos para realizar estos cálculos fueron:

1. Para cada TTL de los *ICMP Echo Request* enviados, a partir del RTT medido, calculamos el dRTT (delta RTT) de cada salto entre cada par de nodos como:

$$dRTT_i = RTT_i - RTT_{i-1}$$

   Donde RTT_i es el valor de RTT en cada paso.

2. Debido a que `traceroute` no es una herramienta del todo precisa, los resultados pueden presentar ciertas anomalías que se manifiestan con resultados extraños o erróneos. Uno de estos comportamientos podría darse cuando, para un valor de TTL el paquete *ICMP Echo Reply* toma un camino más largo que el paquete *ICMP Echo Reply* correspondiente al TTL siguiente. En este caso, el RTT del primero terminará siendo mayor que el siguiente, logrando que el dRTT de este último tome un valor negativo. Esto nunca podría ocurrir en la realidad, ya que el tiempo que tarda un paquete en llegar de un nodo a otro más lejano siempre debería ser positivo.
   Para salvar este comportamiento indeseado y evitar que la búsqueda de outliers se vea afectada, decidimos reemplazar a todos los dRTT negativos por el valor del promedio de todos los dRTT positivos obtenidos. De esta forma buscamos que estos dRTT incorrectos se acerquen más a valores reales.

3. Utilizamos al conjunto de los $dRTT_i$ como los valores de nuestra muestra, con $i = 1, \ldots, n$. Donde n es la cantidad de muestras obtenidas. Ordenamos la muestra de forma creciente.

4. Por cada $dRTT_i$ calculamos el valor absoluto del desvío como:

$$\delta_i = |dRTT_i - \overline{dRTT}|$$

   Donde $\overline{dRTT}$ es el valor de la media de la muestra calculado como el promedio de los $dRTT_i$ medidos.

5. Tomamos como referencia la tabla de valores calculados para la fórmula de *Thompson modificada* del artículo de Cimbala para obtener el valor τ correspondiente a las n muestras. Con este valor obtuvimos:

$$\tau S = \tau * S$$

   Donde S es el desvío estándar calculado como:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(dRTT_i - \overline{dRTT})^2}{n-1}}$$

6. Luego, tomamos el último valor (el máximo) de la muestra ordenada y verificamos, si se cumple que $\delta_i > \tau S$, entonces se asume que el salto con $dRTT_i$ es un enlace intercontinental. Removemos este $dRTT_i$ de la muestra y volvemos a intentar con el nuevo último valor de la muestra hasta que no se cumpla la desigualdad mencionada anteriormente, donde habremos terminado de encontrar los outliers.

De esta forma, mediante una técnica estadística y las herramientas que brinda el protocolo *ICMP*, logramos inferir cuáles son los enlaces intercontinentales en las rutas tomadas por los paquetes.

4. Experimentos y análisis

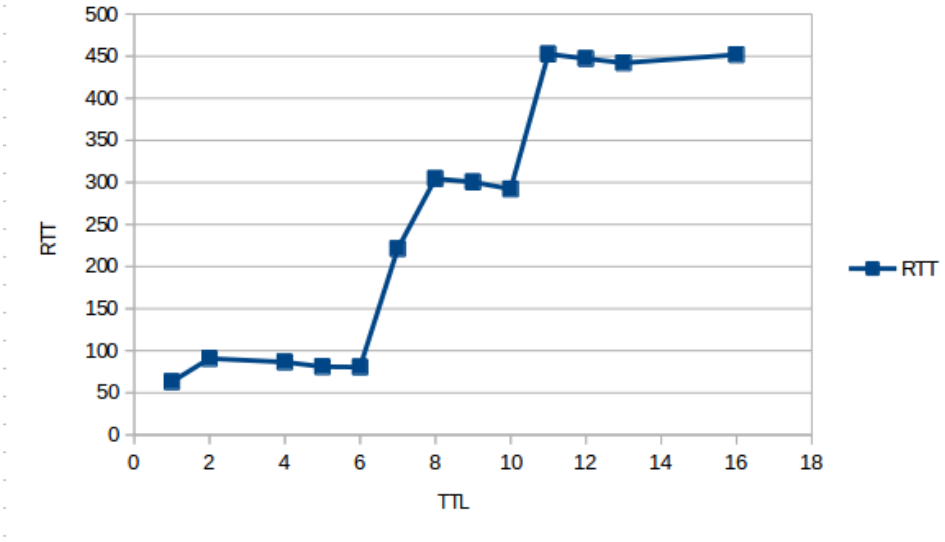
A continuación analizaremos los resultados de algunos experimentos realizados con la herramienta desarrollada.

Para los experimentos elegimos tres universidades de continentes diferentes para analizar las rutas tomadas por los paquetes hacia las mismas. Las universidades que elegimos son:

- Universidad de Sydney, Australia: sydney.edu.au
- Universidad de Ciencia y Tecnología de Trondheim, Noruega: ntnu.edu
- Universidad de Shanghai, China: shu.edu.cn

4.1. Universidad de Australia

Para analizar los resultados obtenidos con la Universidad de Sydney, Australia (sydney.edu.au) primero veremos un gráfico que muestra los valores de los RTT obtenidos para cada salto (TTL).



Se puede observar lo que parecen ser dos grandes saltos, uno entre el TTL 6 y 7, y otro entre el TTL 10 y 11. Estos dos saltos parecerían candidatos a ser enlaces intercontinentales.
Otro detalle que podemos observar es que hay valores de RTT que disminuyen con respecto al anterior, cuando sería de esperar que el RTT aumente en cada salto. Por último, se puede ver que para algunos saltos, no se ven valores de RTT. Esto se debe a que para algunos valores de TTL, el nodo no retornó respuesta. Posiblemente por tener deshabilitado el protocolo ICMP.

A continuación se muestran los resultados obtenidos por nuestra herramienta `traceroute` y realizaremos un análisis más detallado.

Saltos	IP	RTT	DRTT	País
1	192.168.1.1	67.0	67.0	Local
2	201.254.128.1	77.23	10.23	Argentina
3	*****			
4	200.51.208.90	86.1	8.87	Argentina
5	200.51.240.181	89.83	3.73	Argentina
6	213.140.39.118	84.63	-5.2	Argentina
7	176.52.255.27	221.7	137.07	United States
8	213.140.36.70	303.3	81.6	United States
9	213.140.52.229	284.67	-18.63	United States
10	208.185.52.74	300.2	15.53	United States
11	202.158.194.176	443.4	143.2	Australia
12	113.197.15.146	446.6	3.2	Australia
13	138.44.5.47	443.23	-3.37	Australia
14	*****			
15	*****			
16	129.78.5.8	447.28	4.04	Australia

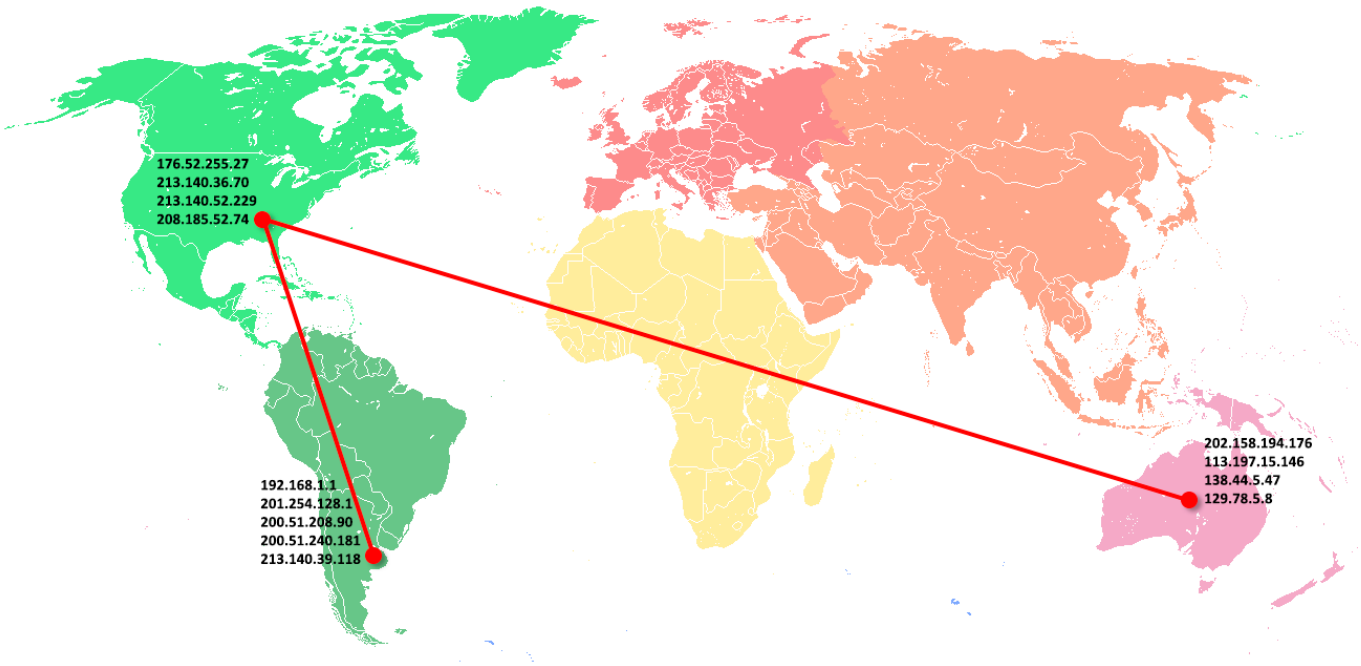
Podemos notar que para los saltos 3, 14 y 15 no se obtuvo respuesta. Esta posiblemente sea una anomalía de `traceroute` conocida como *Missing Hops*. Por lo general ocurre cuando un router está protegido por un firewall o configurado para no generar paquetes *ICMP Time Exceeded*.
También podemos observar algunos valores negativos, en los saltos 6, 9 y 13. Esto puede ser porque se presenta un tipo de anomalía llamada *False Round-Trip Times*. Este caso suele darse cuando los tiempos de ida y vuelta de un paquete reportados por `traceroute` son erróneos. Por lo general puede haber dos razones que generan este comportamiento.
Rutas de paquetes asimétricas o enrutamiento MPLS.
Cuando los respectivos caminos hacia y desde el destino son asimétricos, es decir, que los paquetes se encaminan por caminos diferentes, los tiempos de ida y vuelta pueden no reflejar el RTT real.
Un resultado similar se produce en los enlaces MPLS, donde el paquete tiene que viajar hasta el final de la ruta MPLS, antes de que la respuesta se devuelva al origen. Dado que los routers que manejan exclusivamente MPLS solamente conocen el siguiente salto, no pueden enviar *ICMP Time Exceeded* directamente. En su lugar, tienen que utilizar la ruta por donde el paquete original habría ido. El resultado de esto es, que todos los paquetes viajan hasta el último router MPLS. Por lo tanto, para `traceroute`, los RTT de los saltos en el camino MPLS reflejan aproximadamente el RTT del último router MPLS.

Ahora intentaremos inferir cuáles saltos son enlaces intercontinentales mediante la técnica de estimación de John M. Cimbala. El resultado obtenido por nuestra herramienta fue que los siguientes saltos corresponden a enlaces intercontinentales.

Desde	Hasta	DRTT	País Desde	País Hasta
208.185.52.74	202.158.194.176	143.2	United States	Australia
213.140.39.118	176.52.255.27	137.07	Argentina	United States
176.52.255.27	213.140.36.70	81.6	United States	United States

Como podemos observar, los dos primeros outliers sugeridos por la herramienta parecerían efectivamente ser enlaces intercontinentales.
Para verificar esta hipótesis nos apoyamos en la información provista por sitios de geolocalización de direcciones IP y pudimos comprobar lo siguiente.
El primer salto corresponde a un enlace entre un router de Argentina con uno de Estados Unidos. El segundo, es desde un router de Estados Unidos a uno de Australia.
La herramienta, sugirió un tercer outlier, pero según el análisis de geolocalización que pudimos realizar parecería no representar un enlace intercontinental. Si bien las herramientas de geolocalización ubica al router con dirección IP 176.52.255.27 en España. Pudimos observar que el nombre de este host es hu0-11-0-0-grtmiabr5, donde el texto “mia” corresponde a la abreviación de Miami. Por lo tanto deducimos que corresponde a un router ubicado en Estados Unidos con una dirección IP española asignada.

A continuación veremos en un mapa las direcciones IP de los diferentes saltos, ubicadas según el análisis realizado.

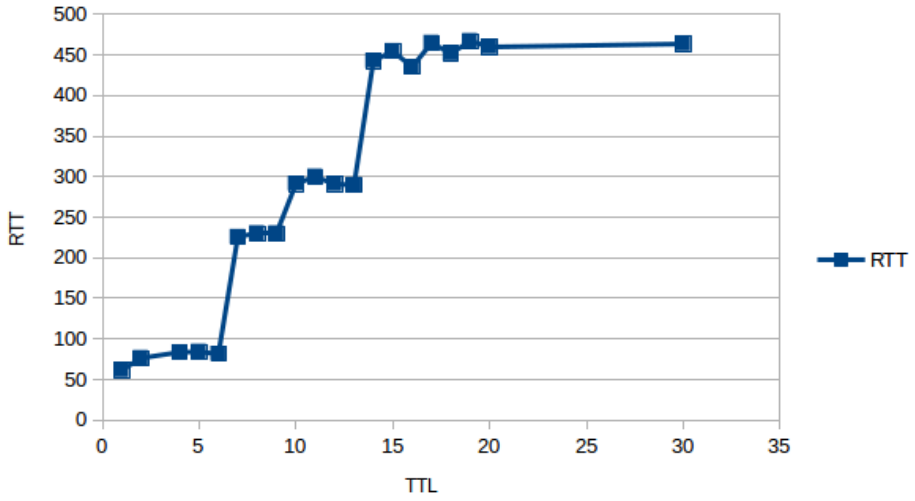


Si bien en el mapa se muestran los enlaces como líneas rectas, para el caso del enlace entre Estados Unidos y

Australia, lo más probable es que este en realidad atraviese el océano Atlántico.

4.2. Universidad de China

En segundo lugar realizamos el análisis de los datos para la Universidad de Shanghai, China (shu.edu.cn). A continuación veremos el gráfico que muestra los valores de los RTT en cada salto (TTL).



En el gráfico, se observan dos grandes saltos en los intervalos 6 a 7 y entre los intervalos 13 a 14. También se destaca otro salto, aunque de menor escala que los anteriores, en el intervalo de 9 a 10. Todos estos saltos parecerían candidatos a ser enlaces intercontinentales. También se nota un gran intervalo sin respuestas entre los saltos desde el 20 al 30. Posiblemente por ser routers que tienen deshabilitadas las respuestas ICMP. Por último, se pueden observar cuatro conjuntos de enlaces con valores de RTT similares.
Vemos a continuación los resultados obtenidos por nuestra herramienta.

Saltos	IP	RTT	DRTT	País
1	192.168.1.1	61.44	-	Local
2	201.254.128.1	76.47	15.02	Argentina
3	*****			
4	200.51.208.90	83.59	7.19	Argentina
5	200.51.240.181	84.23	0.65	Argentina
6	213.140.39.118	81.47	-2.77	Argentina
7	176.52.255.27	225.63	144.17	United States
8	213.140.37.13	230.13	4.5	United States
9	63.243.152.141	229.97	-0.17	United States
10	63.243.152.62	291.2	61.23	United States
11	66.110.72.6	299.93	8.73	United States
12	66.110.57.82	290.7	-9.23	United States
13	66.110.59.182	289.73	-0.97	United States
14	101.4.117.213	442.4	152.67	China
15	101.4.117.97	454.57	12.17	China
16	101.4.112.105	435.1	-19.47	China
17	101.4.112.70	464.63	29.53	China
18	101.4.116.117	451.69	-12.94	China
19	101.4.117.29	466.94	15.25	China
20	101.4.115.173	459.77	-7.17	China
21	*****			
22	*****			
23	*****			
24	*****			
25	*****			
26	*****			
27	*****			
28	*****			
29	*****			
30	202.120.127.220	463.43	3.67	China

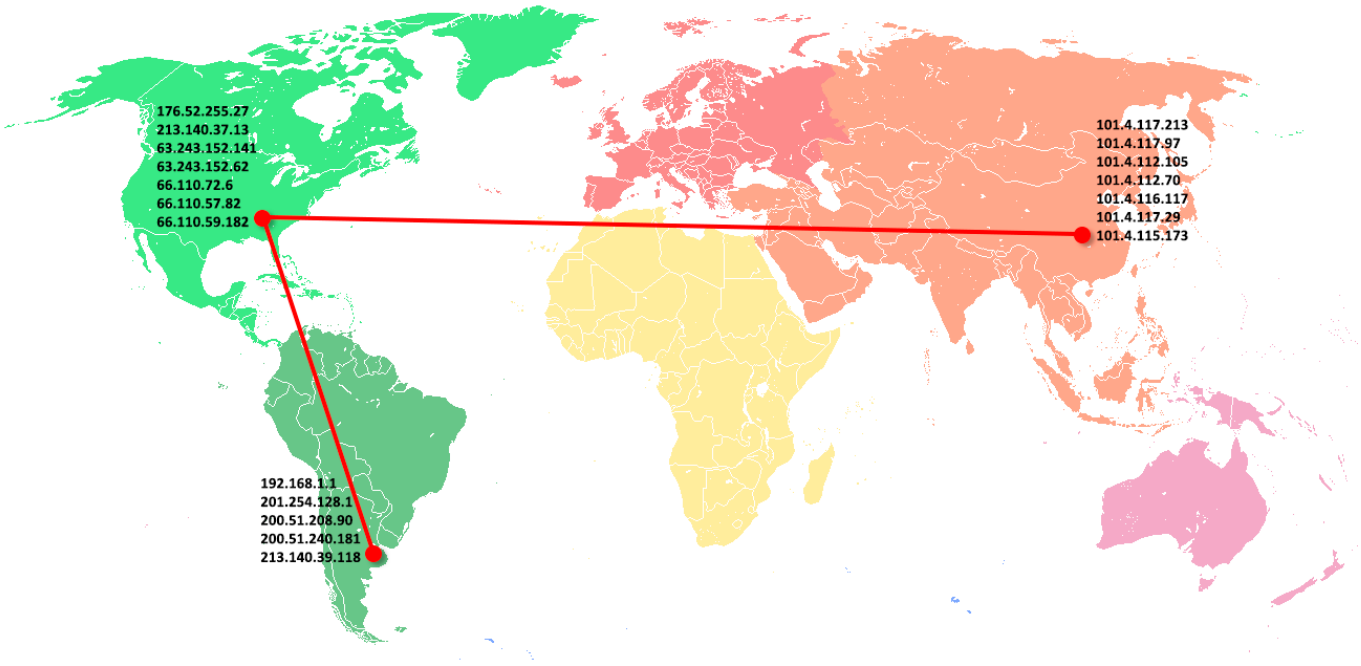
Nuevamente podemos observar algunos saltos en los que no se obtuvo *ICMP Echo Reply*. La mayoría de los routers que no respondieron, corresponden a equipos ubicados en China. Posiblemente esto se deba a la anomalía de `traceroute` llamada *Missing Hops*.
Vemos también que los grupos de enlaces con valores de RTT similares que pudimos observar anteriormente en el gráfico, coinciden con el país en el que se encuentran.
Se vuelve a dar el caso del experimento anterior en el que se calcularon valores de dRTT negativos. Como explicamos anteriormente, este podría ser el caso de la anomalía conocida como *False Round-Trip Times*.

Ahora intentaremos inferir cuáles saltos son enlaces intercontinentales mediante la técnica de estimación de John M. Cimbala. El resultado obtenido por nuestra herramienta fue que los siguientes saltos corresponden a enlaces intercontinentales.

Desde	Hasta	DRTT	País Desde	País Hasta
66.110.59.182	101.4.117.213	152.67	United States	China
213.140.39.118	176.52.255.27	144.17	Argentina	United States
63.243.152.141	63.243.152.62	61.23	United States	United States

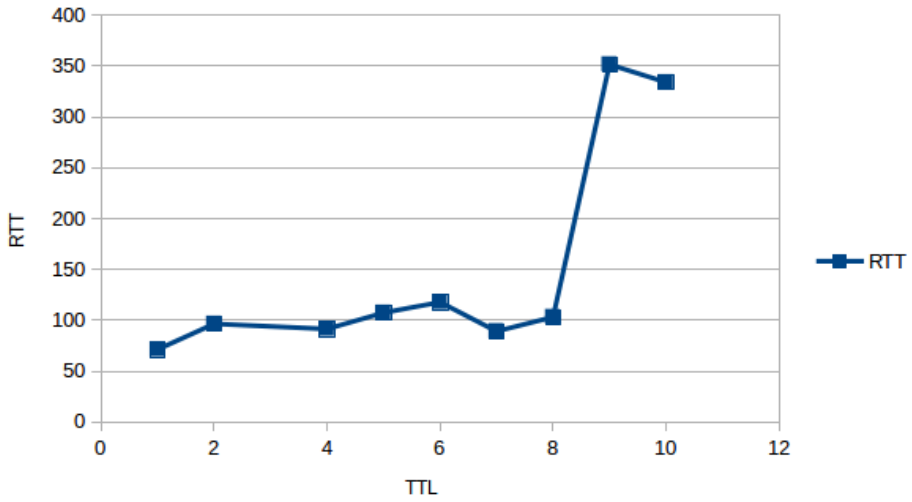
Los resultados que arrojó la herramienta, confirman la hipótesis que los dos grandes saltos observados anteriormente corresponden a enlaces Intercontinentales. De Estados Unidos a China que posee el dRTT más grande y de Argentina a Estados Unidos. El otro salto que observábamos, el test lo calculo como un outlier, pero viendo las ubicaciones de los nodos, esto no sería un enlace intercontinental ya que ambos se encuentran ubicados en Estados Unidos.

A continuación veremos en el mapa las direcciones IP de los diferentes saltos, ubicadas según el análisis realizado.



4.3. Universidad de Noruega

Para el último experimento elegimos una Universidad de un país de Europa. A continuación analizaremos los datos del experimento de la Universidad de Noruega, de Ciencia y Tecnología (ntnu.edu). Veremos el gráfico que muestra los valores de los RTT en cada salto (TTL).



En el gráfico, se pueden observar saltos con similar valor de RTT, pero no todos de manera creciente. Esto posiblemente se deba a diferentes tipos de anomalías del `traceroute`. También podemos ver que para el último salto el valor del RTT es menor que el del RTT del salto anterior.
Además podemos observar un único gran salto que probablemente pertenezca a un enlace intercontinental.

Analizamos los datos obtenidos con nuestra herramienta de traceroute.

Saltos	IP	RTT	DRTT	País
1	192.168.1.1	71.22	71.22	Local
2	201.254.128.1	96.5	25.28	Argentina
3	*****			
4	200.51.208.90	91.47	-5.03	Argentina
5	200.51.240.181	107.4	15.93	Argentina
6	213.140.39.118	117.77	10.37	Argentina
7	213.140.35.83	89.13	-28.63	Argentina
8	213.140.53.77	103.1	13.97	Argentina
9	89.221.43.28	351.67	248.57	UK
10	149.3.183.45	334.17	-17.5	Noruega

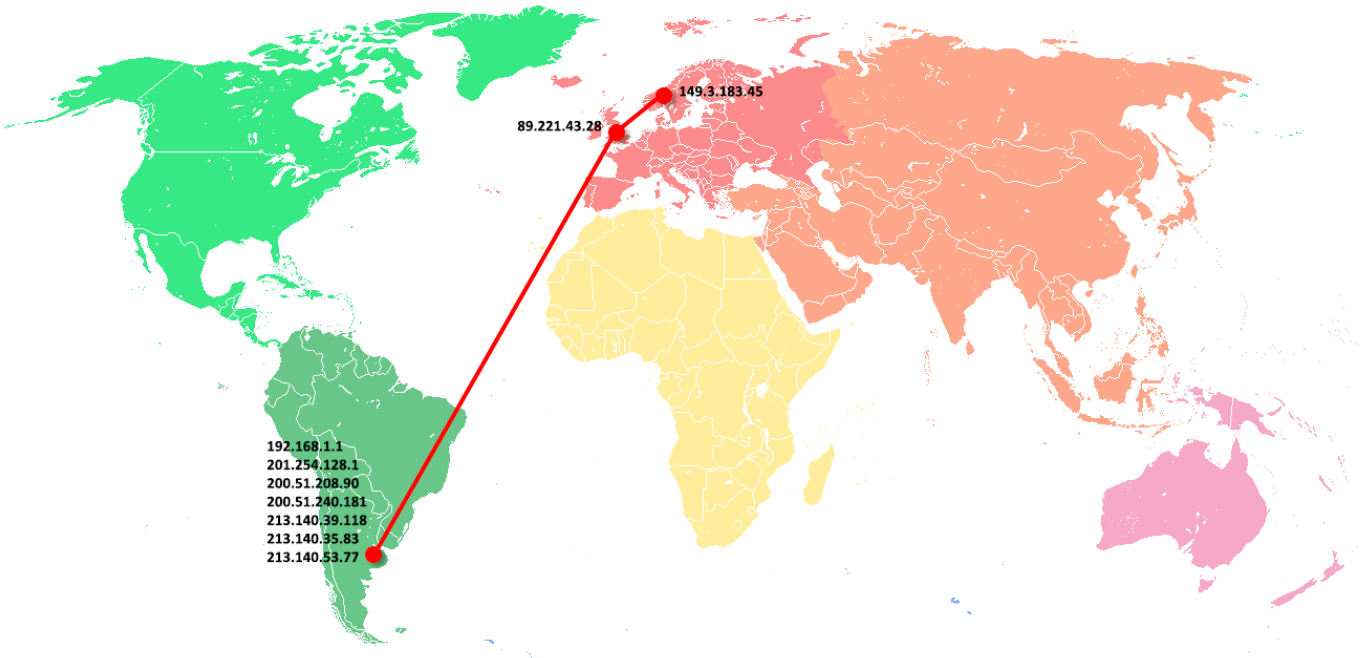
En este caso, se ve solo un salto sin respuesta, posiblemente debido a la anomalía de `traceroute` *Missing Hops*. También se obtuvieron dos enlaces con dRTT negativo, probablemente por la anomalía *False Round-Trip Times*. Por último parecería haber un enlace intercontinental en el salto 9, por su alto valor de dRTT comparado con el resto. Además este nodo se encuentra ubicado en UK y el anterior en Argentina.

Utilizaremos la la técnica de estimación de John M. Cimbala para verificar si efectivamente el salto 9 en un enlace intercontinental. El resultado obtenido por nuestra herramienta fue el siguiente.

Desde	Hasta	DRTT	País Desde	País Hasta
213.140.53.77	89.221.43.28	248.57	Argentina	UK

Efectivamente lo que intuíamos que era un salto intercontinental fue detectado por nuestra herramienta.

A continuación veremos en el mapa las direcciones IP de los diferentes saltos, ubicadas según el análisis realizado.



5. Conclusiones

Para los análisis utilizamos herramientas de geolocalización de direcciones IP. Uno de los principales inconvenientes que tuvimos al usar diferentes herramientas de este tipo fue que no siempre coinciden en la respuesta. Pero, gracias a la interpretación de los RTT en cada salto, los nombres de los hosts y la ubicación de los hosts vecinos, pudimos inferir cuál sería su verdadera ubicación.

Fue muy frecuente, en las pruebas realizadas, que el primer gran salto se de a un nodo ubicado en Estados Unidos, como se vio en los resultados para los casos de las Universidades de Australia y China.

La cantidad de saltos observados en las pruebas realizadas, siempre fue inferior a 30 para aquellas pruebas que terminaron correctamente. Cuando el número fue mayor a 30, no se llegó a recibir respuesta del destino.

También, nos encontramos con nodos intermedios de los que no tuvimos respuesta. Esto como se explicó anteriormente suele darse por la anomalia de traceroute llamada *Missing Hops*.

Otra cosa que pudimos observar es que para algunos saltos se mostraban RTTs menores que RTTs de saltos anteriores que se explica por la anomalía llamada *False Round-Trip Times*.

Otra característica que pudimos identificar fue que en varios casos existen hosts con una dirección IP asignada correspondiente a al rango de direcciones de otro continente. En este caso las herramientas de localización geográfica ubicaban a estos hosts en el continente correspondiente según su IP en lugar de proveer su ubicación real.

Este trabajo nos permitió conocer los detalles sobre la implementación de la herramienta `traceroute` y el funcionamiento del protocolo ICMP. También pudimos realizar análisis sobre los resultados obtenidos al utilizar nuestra propia implementación de traceroute en diferentes experimentos. Con esto pudimos verificar que los resultados obtenidos no siempre fueron exactos, sino que requirieron de cierto análisis para su interpretación.