DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 2

Taller de Traceroute

4 de mayo de 2024 — Teoría de las Comunicaciones

Integrante	LU	Correo electrónico
Alexandra Abbate	710/19	alexandra-abbate@hotmail.com
Lucas Mas Roca	122/20	lmasroca@gmail.com
Nicolás Valentini	86/21	nicolasvalentini@hotmail.com
Alan Roy Yacar	174/21	alanroyyacar@gmail.com

Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria - Pabellón I
Intendente Güiraldes 2160 - C1428EGA
Ciudad Autónoma de Buenos Aires - Argentina
Tel/Fax: (54 11) 4576-3359
http://exactas.uba.ar

1. Introducción

El eje central de este trabajo es poder analizar el recorrido que realiza un paquete IP a través de internet. Se va a hacer un enfoque en detección de saltos interoceánicos. La experimentación de este trabajo consta del envío de paquetes IP encapsulados en paquetes ICMP con destino a distintas universidades ubicadas en lugares geográficamente muy distintos.

Para realizar este experimento se van a enviar múltiples paquetes IP encapsulados en ICMP con distintos TTL a las universidades. Luego se calcula el RTT de estos paquetes cuando vuelven con Time Exceeded y con la IP de origen de este último paquete. Con esta IP se puede ubicar la posición geográfica del lugar al que se llegó con una cantidad determinada de saltos. La idea es detectar un significativo aumento del RTT entre dos TTL para determinar que en esa instancia ocurrió un salto interoceánico.

2. Métodos y condiciones de los experimentos

Para realizar los experimentos, se expandió sobre el código provisto por la cátedra. En la implementación, se aumentó a 30 veces el envío de paquetes con TTL de 1 a 25. Lo que espera con esto es poder tener resultados representativos de los RTT que tienen cada uno de los paquetes con los diferentes TTL, de esta forma reduciendo el peso de casos límites. Una vez obtenidos estos resultados, se determina cuál es la IP que más veces aparece para cada TTL, y se saca un promedio de los RTT entre los 30 paquetes enviados para el mismo TTL. Con estos datos, se puede determinar también cuán grande fue el salto de RTT entre dos TTL. Una vez determinados los saltos más grandes en cuestión de RTT, estos quedan como posibles candidatos para el salto interoceánico. Luego, estas hipótesis fueron verificadas utilizando las herramientas provistas por la cátedra.

Adicionalmente, se utilizó el paquete requests de python para obtener las coordenadas de una IP utilizando las distintas páginas de geolocalización de IPs provistas por la cátedra.

Se tomaron cuatro rutas distintas para realizar la experimentación, de las cuales se tomaron distintas universidades como destino en diferentes continentes, estos destinos son:

- La página de la Univerisdad de Sudafrica (UNISA): www.unisa.ac.za

  Esta medición la hicimos a las 6pm un día domingo.

- La página de la Humboldt University of Berlin en Alemania: www.hu-berlin.de

  Este traceroute se realizó a las 10pm un día sábado.

- La página de la Universidad de Pekín en China: www.english.pku.edu.cn

  Esta ruta fué un día domingo a las 3pm.

- La página de la Universidad de Los Angeles, California (UCLA): www.www.ucla.edu

  Este traceroute se realizó a las 10pm un día domingo.

Por último, fueron realizados los mismos experimentos desde la Facultad de Exactas y Naturales, con la intención de analizar el comportamiento del traceroute desde una red con tecnología diferente. Estos experimentos fueron realizados un día lunes a las 16 horas.

3. Resultados

3.1. Universidad de California (UCLA)

Con los resultados obtenidos, se realizó una progresión entre los saltos de TTL y la evolución del RTT total. En este caso, con un TTL de 17 ya se obtuvo una respuesta del destino con 150ms de RTT total. La longitud de la ruta (tomando en cuenta únicamente los saltos que respondieron) es de 12 saltos, respondieron un 63 % de los nodos.

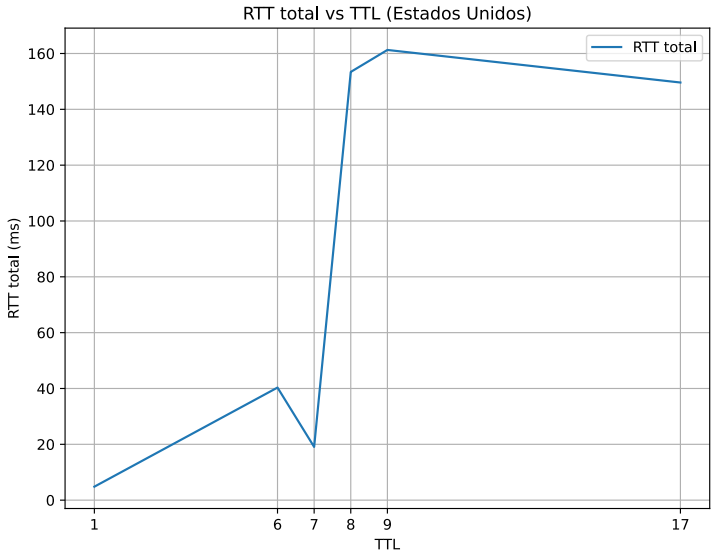

Figura 1: Gráfico de evolución RTT totales desde domicilio hasta Estados Unidos

Como se puede observar en la Figura 1, queda fuertemente marcado que entre el TTL 7 y 8 hubo un salto significativo respecto al RTT total. Esto puede ser un indicador clave de que en ese momento ocurrió un salto geográfico grande. Esto luego fue confirmado, verificando que en el TTL 7, la IP correspondía al nodo 195.22.220.56 ubicado en Buenos Aires, y luego el TTL 8 era de 89.221.41.161 ubicada en Miami, Estados Unidos. En este resultado se pueden detectar múltiples anomalías. Para empezar, de TTL 2 a 5 no se obtienen respuestas, este es un caso de la anomalía de "Missing Hops" donde hay nodos que no generan el error TTL exceeded, esto se puede deber a un firewall o que fue configurado de esa forma, lo mismo sucede del 10 al 16.

Otra anomalía visible es la de "False Round-Trip Times" más comúnmente se nota cuando se encuentran tiempos de RTT entre saltos con una diferencia significativamente negativa, en este caso se puede ver esta anomalía entre el TTL 6 y 7. Esto usualmente se debe a rutas con caminos de ida y vuelta asimétricos, en este caso el dispositivo 6 tiene una ruta de vuelta más lenta que la ruta de vuelta utilizada por el dispositivo 7, causando que se reciban más rápido las respuestas de esta última.

3.2. Universidad de Sudáfrica (UNISA)

Para este experimento, la respuesta se obtuvo con un TTL de 18 con 320ms de RTT para este salto. La longitud de la ruta es de 13 saltos, respondieron un 72 % de los nodos.

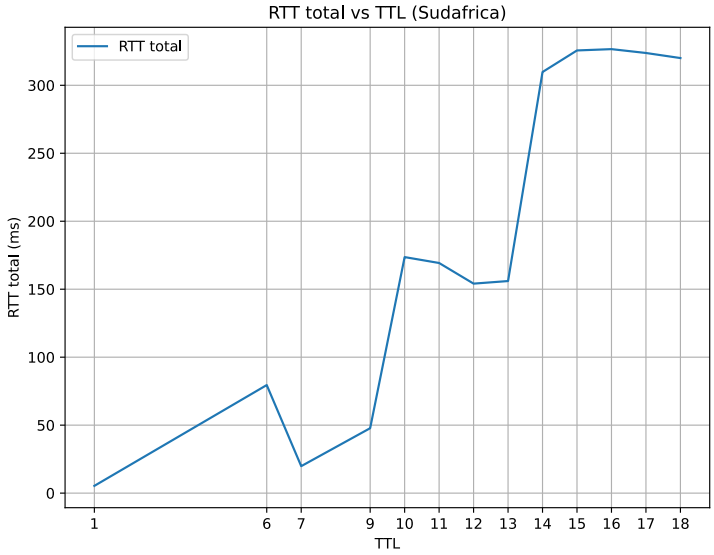

Figura 2: Gráfico de evolución RTT totales desde domicilio hasta Sudáfrica

Siguiendo lo obtenido en la Figura 2, los principales candidatos para saltos intercontinentales, parecerían ser claramente los que van de TTL 9 a 10, y de 13 a 14, ya que presentan una diferencia positiva de RTT notable. Esto fue luego corroborado, viendo que el TTL del 9 correspondía a la IP 149.3.181.65 ubicada en Brasil y el TTL 10 es 129.250.2.196 ubicada en Estados unidos. Un resultado no esperado fue el salto de 13 a 14, ya que ambas IPs estaban ubicadas dentro de Estados Unidos, estas IPs son 170.39.8.30 y 190.103.185.187 respectivamente. Fue recién con TTL 15 que se obtuvo una IP dentro de Sudáfrica, siendo la IP 155.232.1.149. Hay dos posibles explicaciones para este fenómeno, una es simplemente que con las herramientas provistas no fueron suficientes para ubicar correctamente la posición geográfica de las IPs. Otra posible explicación es que debido a la anomalía explicada anteriormente sobre false round-trip times. Manteniendo la línea de RTTs falsos, también se puede ver con el TTL 6. Por último, hay algunos missing hops, estos son fácilmente detectables, entre TTL 1 a 6 y el TTL 8.

3.3. Universidad Humblodt, Alemania

En este caso, la respuesta se obtuvo con un TTL de 19 con 260ms de RTT. La longitud de la ruta es de 12 saltos, respondieron un 63 % de los nodos.

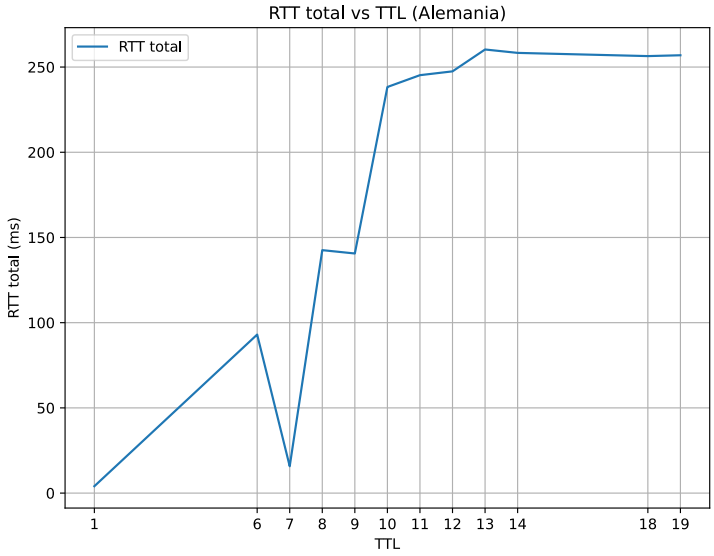

Figura 3: Gráfico de evolución RTT totales desde domicilio hasta Alemania

Observando la Figura 3, sin dudas este es el ejemplo más claro de false Round-Trip Time, marcando muy en claro el salto entre el TTL 1 al 6, siendo que parece como un mayor candidato de salto intercontinental el salto del TTL 7 al 8 y del 9 al 10, mirando la diferencia positiva de RTT entre ambos. Verificando los resultados obtenidos con las herramientas provistas, se pudo ver que la TTL 7 era la IP 195.22.220.56, la cual ya había aparecido también para la de UCLA, está en Buenos Aires y la IP 195.22.199.65 para el TTL 8, la cual está ubicada en Estados Unidos. Por otro lado, la IP 4.68.62.57 del TTL 9 está también ubicada en Estados Unidos, y la IP 4.69.167.114 del TTL 10 está en Alemania, probando así la teoría de que en ambos saltos hubo un salto intercontinental. Por último, se puede ver claramente que entre TTL 1 y 6 hay missing hops y también hay de 14 a 18.

3.4. Universidad de Pekín, China

Para este experimento, se obtuvo una respuesta tras los paquetes con TTL 24, con un RTT promedio de 375ms. La longitud de la ruta es de 19 saltos, respondieron un 79 % de los nodos.

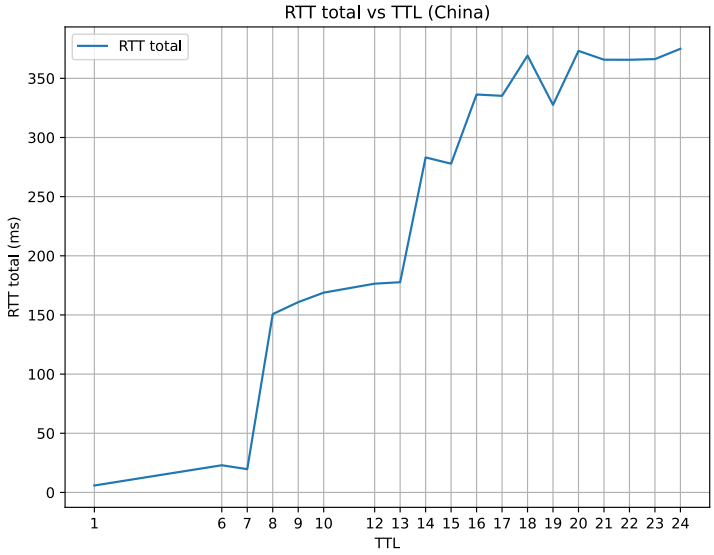

Figura 4: Gráfico de evolución RTT totales desde domicilio hasta China

En primer lugar, este experimento fue un poco más complicado de realizar debido a que múltiples universidades en Asia fueron experimentadas antes que esta, pero ninguna terminaba dando respuesta, este comportamiento tampoco cambiaba con TTLs hasta 30. Esta anomalía puede atribuirse a missing destination y una posible explicación es que el destino sea protegido por un firewall.

Sin embargo, una universidad asiática que dio respuesta fue la Universidad de Pekín. La cual en sus resultados posee múltiples posibles saltos intercontinentales candidatos. Estos podrían ser de 7 a 8, de 13 a 14 y hasta de 15 a 16. Comprobando estas hipótesis, la primera fue del TTL 7 con IP 195.22.220.56 a TTL 8 con IP 195.22.199.82, y efectivamente estas IPs describen un salto geográfico de Argentina a Estados Unidos. Luego hay como candidato del TTL 13 con IP 129.250.3.219 al TTL 14 129.250.3.192, este es un salto de Estados Unidos a Japón, haciendo otro salto intercontinental. Ya en menor escala está el salto de 129.250.5.23 correspondiente al TTL 15 hacia la IP 129.250.2.51 del TTL 16. Este es salto representa un salto de Japón a China, no tan amplio como los otros, pero un salto notable. Por fuera de las anomalías ya notadas antes de missing hop de 1 a 6 y de 10 a 12 y ciertos menores false round-trip times, hay una anomalía en particular que llama la atención, y es el RTT total de los paquetes con TTL 18, 19, y 20. Estos paquetes están todos ubicados dentro de China, sin embargo los 18 y 20 pasan por Hong Kong y el 19 termina en Beijing. Hay dos posibles explicaciones para este fenómeno, uno es simplemente un caso más de false round trip time. Otra hipótesis puede ser simplemente una varianza estadística debido al alto número de hops, y a los altos números de RTT con los que se está trabajando.

3.5. Traceroute desde Facultad de Exactas

Como ultima experimentación se realizaron los mismos envíos desde la Facultad de Exactas y Naturales. Se obtuvieron algunos resultados un tanto diferentes, principalmente en la longitud del recorrido y el porcentaje de respuesta. Ambos aumentaron significativamente. Por otro lado, se detectó un aumento de los RTT totales y el salto intercontinental pareciera ocurrir un TTL antes que los anteriores, ya que en la gran mayoría ocurría en el TTL 7, y desde la facultad en el TTL 6. Para visualizar un resultado representativo de los obtenidos se puede observar la figura 5.

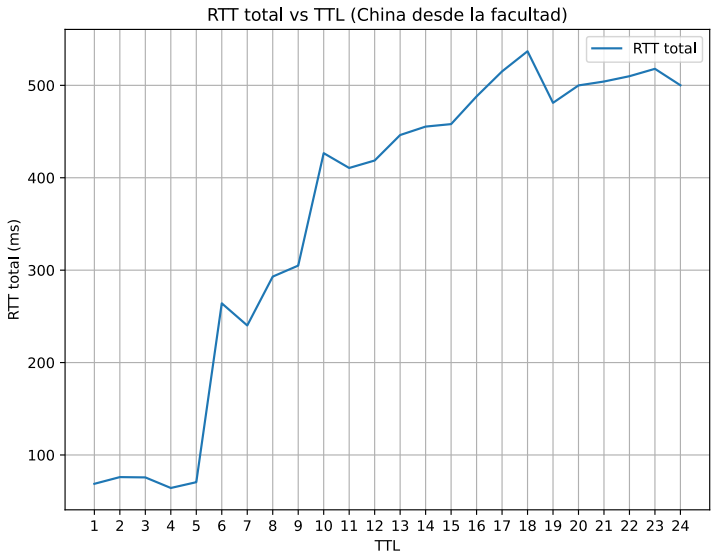

Figura 5: Gráfico de evolución RTT totales desde facultad hasta China

En este caso la longitud de la ruta es de 24 saltos y hay 100 % de respuesta los nodos. Queda claro que los candidatos a saltos intercontinentales son de TTL 5 a 6 y de 9 a 10. En el caso de 5 a 6, el salto es de IP 170.51.254.174 ubicada en la Argentina hacia la IP 168.143.228.100 en Estados Unidos. Por otro lado, del 9 al 10 es desde la IP 129.250.3.219 en Estados Unidos, hacia la IP 129.250.3.192 en Japón. Puede notarse que este salto es el mismo que en el experimento anterior. Finalmente, se puede notar que el RTT es significativamente más alto que en el experimentado desde la red privada. Desde este resultado, se puede detectar un muy bajo número de anomalías.

Una anomalía observada fue la aparición de loops cuando se realizó el experimento en la facultad hacia UCLA. Estos ciclaban con los últimos 3 nodos, entre los que estaba ubicado el nodo destino. Una posible explicación para esto es el manejo de balanceo de cargas en la red, lo cual termina generando un ciclo entre los nodos.

4. Conclusiones

Como conclusión, se puede notar que tomar al traceroute desde la facultad de ciencias exactas y naturales se consigue una mayor consistencia en la cantidad de respuestas obtenidas de cada salto en comparación a una red privada. Esto es especialmente evidente al ver que en todos los experimentos realizados en la red domiciliaria se puede ver que los nodos desde el 2 al 5 nunca responden, mientras que desde la facultad todos estos responden, más aún podemos ver que la mayoría (aproximadamente 69 %) de estas direcciones ip se encuentran dentro del rango de direcciones ip privadas (10.0.0.0/8).

Entre todos los experimentos se pudo ver que en la gran mayoría, resultó una buena medida tomar como potencial salto intercontinental a los que representan un aumento significativo y constante del RTT total. Esto fue válido para casi todos los experimentos, excepto para el de Sudáfrica, que presento algunas anomalías que también aparecieron en otros experimentos.

En término de anomalías, se detectaron múltiples. En todos los experimentos se encontraron missing hops, los cuales redujeron significativamente la longitud de la ruta y el porcentaje de respuesta de los nodos. Por otro lado, la otra anomalía más común fue false round-trip time, el cual se registraba comúnmente antes del primer salto interoceánico y también a lo largo del camino. En el caso de Sudáfrica, esta anomalía presentó una contaminación significativa de los datos, ya que debido a este false round-trip time, parecía como si se adelantara el salto interoceánico. Otra anomalía encontrada fue missing destination, la cual fue especialmente común al enviar paquetes a Asia. Finalmente, fue encontrada la anomalía de loop and circles en el destino del envío del paquete hacia UCLA desde la red de Exactas.

Por otro lado, hubo anomalías que no fueron detectadas, posiblemente debido a que la metodología de los experimentos no era capaz de detectarlas. Entre ellas se encuentran missing links, false links y diamonds. Todas estas anomalías requieren un conocimiento más específico de la topología de la red, y la experimentación realizada, que consistió en observar los round-trip times totales promediados con las IPs más comunes para cada TTL, no permite observar múltiples caminos para un mismo paquete.

En términos generales, concluimos que todos estos errores y anomalías hacen que el uso de esta herramienta carece de utilidad y practicidad para entender las características de la ruta de la red.

5. Anexo: Detección automática de saltos interoceánicos

5.1. Metodología

Para este experimento primero se implementó el método de detección de outliers de Cimbala, utilizando como distribución las muestras positivas de RTT entre saltos. Luego se comparó ese criterio con usar un valor de corte fijo para $\frac{x_i - \bar{x}}{S}$ (siendo x_i las muestras de la distribución, $\bar{x}$ la media y S el desvío muestral). Para comparar los criterios se evalúa la cantidad de falsos positivos y falsos negativos de ambos criterios. En la gran mayoría de los experimentos realizados, las distribuciones muestran outliers según el método de Cimbala, sin embargo, se pueden detectar tanto falsos positivos como falsos negativos. En particular, parecería que en las rutas más largas el método de Cimbala presenta más falsos positivos, mientras que en rutas más cortas parece ser más propenso a falsos negativos.

Luego se experimentó con valores de corte fijo, para esto se consideró primero establecer 1 como valor de corte fijo, ya que $\frac{x_i - \bar{x}}{S}$ es la normalización de la distribución de los RTT entre saltos, con lo cual si se toma como valor de corte fijo 1 entonces un x_i se considera como outlier si y solo si:

$$\frac{x_i - \bar{x}}{S} > 1$$

$$x_i - \bar{x} > S$$

con lo cual un dato es considerado como outlier si se aleja positivamente de la media por una cantidad mayor que el desvío muestral. Tomando este criterio, se obtienen mejores resultados que los del método de Cimbala, se encuentran menos falsos positivos y falsos negativos.

Luego, aprovechando que estamos definiendo un criterio de corte sobre una normalización de una distribución, la cual se sabe que en principio debería comportarse de forma similar a una distribución normal estándar, podemos tomar un percentil de la distribución normal estándar y definir su valor como corte fijo. En este caso se tomó el percentil 80 que en la distribución normal estándar es aproximadamente 0.842 y se usó como criterio de corte fijo (si $\frac{x_i - \bar{x}}{S} > P_{80} \approx 0{,}842$ entonces x_i es considerado outlier). Se puede ver como hacer el criterio más laxo a outliers de esta forma permite una mayor detección de saltos interoceánicos, mejorando la precisión del método aún más. Sin embargo, vale la pena notar que si se sigue disminuyendo este valor será cada vez más probable caer en falsos positivos.

A continuación se presentan algunos de los experimentos con resultados más relevantes y representativos.

5.2. Outliers de Cimbala y corte fijo en China

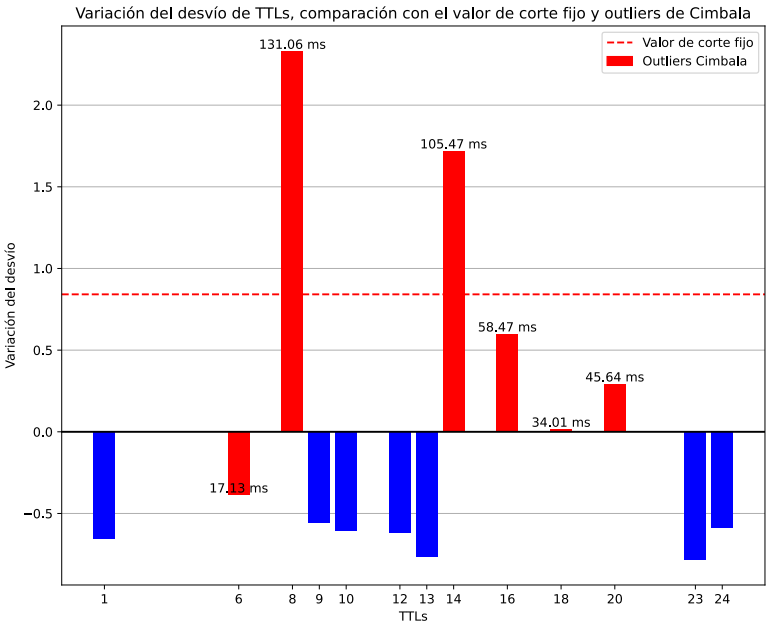

Figura 6: Outliers de China desde domicilio

A partir de la figura 6 se pueden observar múltiples fenómenos. Pare empezar el camino fue de los más largos obtenidos y esto pareció afectar significativamente la detección de outliers según Cimbala. Un dato ciertamente cuestionable de Cimbala fue la elección como outlier a un salto cuyo RTT hop fue menor a la media entre los RTT hops positivos. Lo que ocurrió en este caso fue que Cimbala no detecto ningún falso negativo, pero sí múltiples falsos positivos. Esto se podría explicar, ya que Cimbala es un método de detección de outliers dentro de una distribución, y no es precisamente lo buscado en este proceso, por ende, datos anómalos detectados por Cimbala que serían outliers por debajo de la media, no tiene mucho sentido para este trabajo que sean detectados. Por el contrario, si se observan los valores y se los compara con el valor de corte fijo, los únicos outliers son los saltos interoceánicos, con lo cual con esta medida no hay ni falsos positivos ni falsos negativos.

5.3. Outliers de Cimbala y corte fijo en Alemania

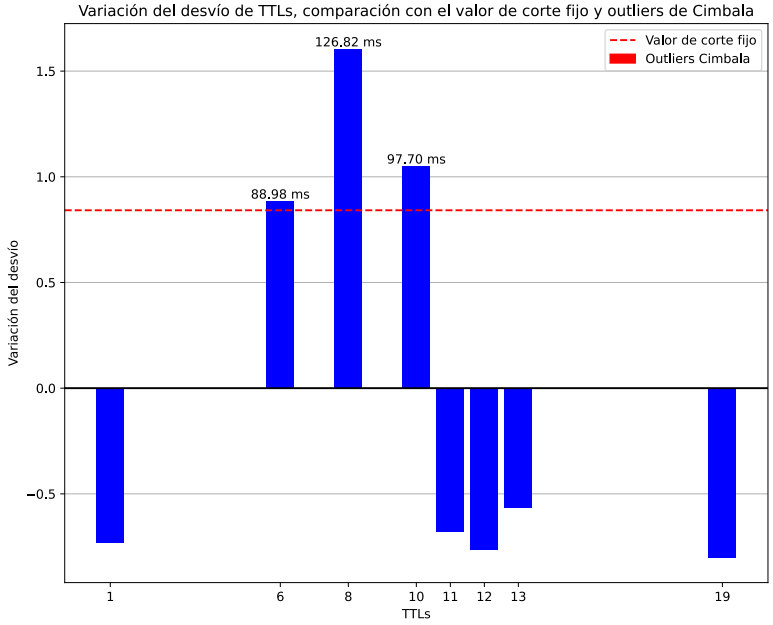

Figura 7: Outliers de Alemania desde domicilio

En la figura 7 podemos observar como Cimbala no detecto ningún outlier, con esto claramente no está detectando los saltos interoceánicos que debería detectar. Estos falsos negativos pueden deberse a que por la forma en que se van tomando los datos, Cimbala calcula el corte a partir de un tau que se calcula de forma dinámica y por ende descarta casos que deberían serlos según el contexto. También podría pasar que no tome outliers, ya que los saltos interoceánicos por mucho que tengan RTT hop altos no sean tan significativamente distintos a la media como para ser detectados como outliers. Un caso contrario es el de corte fijo, en donde de hecho ocurre que se detecta un falso positivo. El falso positivo corresponde al del TTL 6 el cual es un salto dentro de la Argentina, una posible explicación de este fenómeno es que del nodo 1 al 6 no se detecta ningún hop, por lo tanto, el salto al 6 puede resultar en uno muy notable, logrando así estar como candidato a interoceánico cuando los RTT hops no son tan grandes. Finalmente, en caso de haber tomado como corte fijo 1, este falso positivo no se hubiese dado.

5.4. Outliers de Cimbala y corte fijo en Sudáfrica

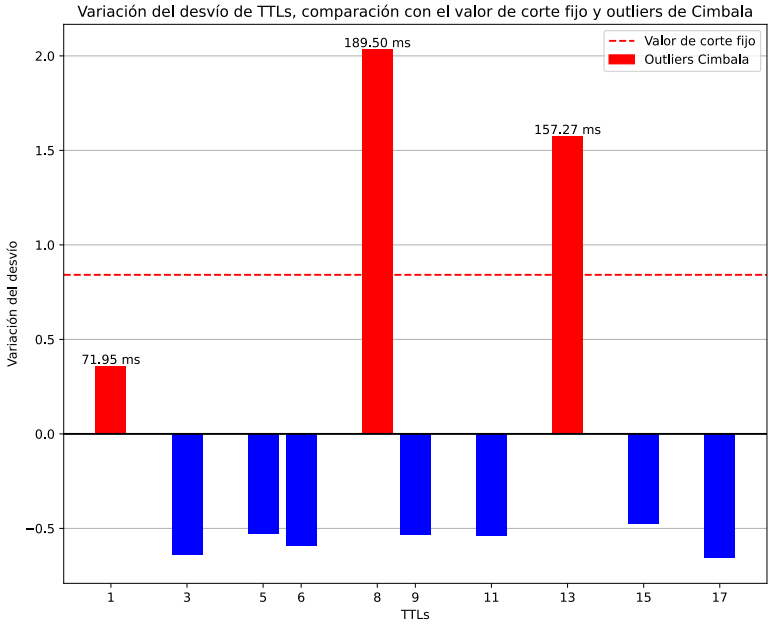

Figura 8: Outliers de Sudáfrica desde la facultad

Finalmente, en la figura 8 se puede notar que ambos métodos dieron resultados parecidos entre sí, en donde ninguno dio falsos negativos, Cimbala obtuvo dos falsos positivos y el de corte fijo un falso positivo. Algo importante a destacar de esta muestra es este experimento presentaba el mismo problema que en el experimento de la sección 3.2, esto quiere decir que justo antes del salto interoceánico verdadero, hay un false round-trip time muy alto que contamina el RTT hop del siguiente, y por ende termina haciendo básicamente imposible detectar ese salto interoceánico. Por ende, ambos registraron erróneamente el TTL 13, y en el caso de Cimbala también el TTL 1, el cual fue un salto dentro de la misma facultad, sin embargo es cierto que fue un salto significativo en términos de tiempo.

5.5. Concluciones

De estos experimentos se puede concluir que el método Cimbala puede ser muy práctico en una gran variedad de casos, sin embargo, dentro del área de detección de saltos interoceánicos en un traceroute, puede llegar a ser ineficiente en comparación a otros métodos. Principalmente detectando un alto número de falsos positivos en caminos de mayor largo, y un alto número de falsos negativos en caminos cortos. Cuando se trataba de caminos con la varianza positiva fuertemente marcada tuvo mejores resultados, sin embargo, en todos los casos fueron peores que los del corte fijo. Por otro lado, se vio que dentro de corte fijo, si se hubiese tomado como corte el 1, hubiese habido un falso positivo menos en los experimentos, sin embargo, mantenemos que este número puede llegar a ser muy restrictivo en muchas redes y caminos, y puede llegar a ser más representativo tomar el percentil 80 de la distribución normal estándar. Por último, se puede observar que sin importar el método, hay ciertos fenómenos que le fueron imposibles de detectar debido al comportamiento de la red, como fue en el caso de Sudáfrica.