Detección de Enlaces Intercontinentales

Manuel Costa Mathias Gatti

Resumen —

Index Terms —**traceroute, enlaces intercontinentales, ICMP, anomalías, RTT, TTL**

I. Introducción

— COMPLETAR —

Iniciamos este trabajo con el objetivo de aprender sobre el protocolo ARP y más en general las redes de Internet y sus algoritmos de intercambio de paquetes.

A lo largo de este informe describiremos ciertos análisis que tendrán como objetivo el modelado de emisores de información y la posterior detección de símbolos destacados. A partir de esta metodología creemos que podremos identificar dispositivos claves en la red como por ejemplo el gateway. Nuestra hipótesis es que este tendrá destacará en el intercambio de paquetes who-has siendo de los que mas reciba y envíe.

— COMPLETAR —

II. Métodos

Herramientas

Para la implementación de traceroute utilizamos el código provisto por la catedra al cual le realizamos ciertas modificaciones para poder detectar anomalías y guardar los datos obtenidos de forma más cómoda.

Detección de saltos intercontinentales

Para la detección automática de saltos intercontinentales aplicamos una técnica basada en el Modified Thompson Tau Test para detección de outliers, como se explica en Cimbala[1]. Dicho test consiste en comparar el valor absoluto de las muestras estandarizadas (mediante z-score) contra un estadístico, τ, que depende del tamaño de la muestra. En particular, nuestra versión difiere con la presentada con la de Cimbala en que no tomamos el valor absoluto del z-score, dado que no estamos interesados en detectar los casos atípicamente pequeños.

Asunciones realizadas

- A los fines prácticos, vamos a considerar que un salto de América del Sur a América del Norte se considera un salto intercontinental, por la extensión del mismo.
- A veces se da el caso en que el RTT promedio para un cierto TTL puede ser menor que el RTT del TTL anterior. Ante esta situación seteamos el RTT diferencial (delta) en 0. Razones por las que puede suceder esto son el problema de los *caminos asimétricos*, o bien que haya un desvío estándar elevado y el hop realizado sea corto. Se nos presento entonces la duda de si considerar estos valores o no a la hora de hacer el cálculo de los *outliers*. Decidimos que tanto un problema como el otro pueden estar afectando a otros hops que sin embargo no llegaron a dar 0, pero dieron un valor menor al que deberían. Por lo tanto, sería injusto (y posiblemente un error metodológico) solo omitir a los valores nulos.

Rutas

Se corrió traceroute sobre 3 universidades distintas con un ttl de 30 y 40 queries. Con el objetivo de lograr contrastar elegimos universidades muy lejanas y muy cercanas. A continuación describimos brevemente a cada una.

- Universidad de São Paulo (www5.usp.br) esta será la universidad mas cercana, ubicada en el mismo continente, por lo cual esperamos que no haya ningún enlace intercontinental.
- Universidad de Sidney (www.sydney.edu.au) escogimos esta universidad ya que nos surgió la duda de si existirá algún enlace intercontinental directo entre oceanía y america o tendrá que pasar europa resultando en varios enlaces.
- Universidad de Moscú (www.msu.ru) al estar ubicada en un punto tan alejado de nosotros estabamos seguros de que iba a haber algún salto intercontinental y quizás más.

III. Resultados y análisis

Los resultados de cada ruta se presentan en dos partes: primero se realiza una descripción geográfica de la ruta trazada, basada en las ips seguidas, prestando particular atención a los saltos intercontinentales; en la segunda parte se analizan los RTTs entre saltos, de los cuales se busca poder inferir en forma automática los enlaces intercontinentales detectados en la primer parte.

Universidad de São Paulo

Recorrido en el Planisferio

COMPLETAR: En esta parte creo que deberiamos responder las preguntas ¿Que porcentaje de saltos no responden los Time exceeded? ¿Cual es el largo de la ruta en terminos de los saltos que si responden? (todo esto se puede sacar viendo los csv) ¿La ruta tiene enlaces intercontinentales? ¿Cuantos?

M. Costa, e-mail: manucos94@gmail.com
M. Gatti, e-mail: mathigatti@gmail.com

[1] http://www.net.in.tum.de/fileadmin/TUM/NET/NET-2012-08-1/ NET-2012-08-1_02.pdf

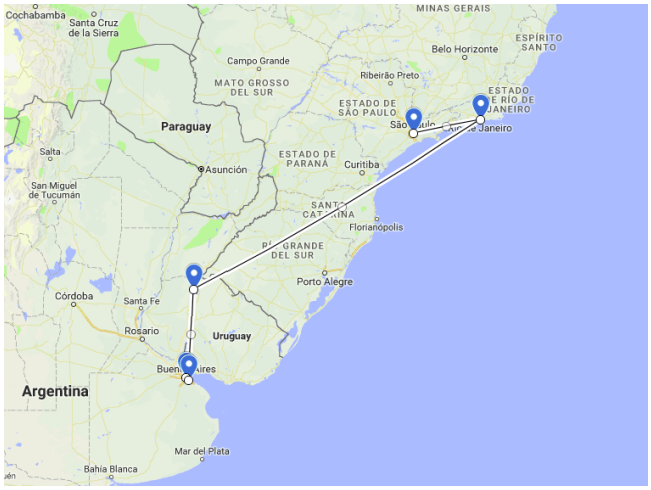

Figura 1: Recorrido realizado por los paquetes durante la ejecución de traceroute al intentar alcanzar el sitio `www5.usp.br`

RTT entre saltos

Analicemos cómo funciona nuestro modelo para inferir saltos intercontinentales, en un caso donde sabemos que no hay ninguno.

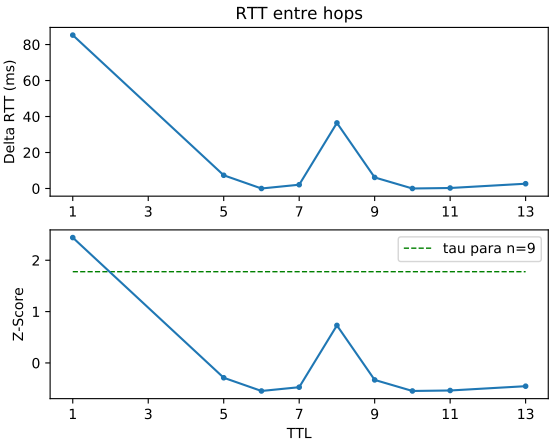

Figura 2: RTT entre saltos (antes y después de normalizar respectivamente) para el sitio `www5.usp.br`. n es la cantidad de TTLs para los que se obtuvo un *time exceed* (que son los puntos que se grafican).

Viendo la figura 2, observamos que hay un falso positivo: el RTT estandarizado del primer hop supera el umbral dado por la τ del Test de Thompson. Podemos entender esto como una consecuencia de que el test utilizado no hace más que buscar valores atípicos (en nuestro caso atípicamente grandes) dentro de una muestra. Por lo tanto, en la medida que los RTTs entre hops no sean equitativos, no es de sorprender que casi siempre encontremos algún salto que sobresalga del resto. La clave está en que si tuviéramos un trayecto significativamente más largo, este valor que ahora resulta un outlier muy posiblemente quedaría opacado por el RTT diferencial de un verdadero salto continental.

Vale decir que en este caso particular, lo que parece estar sucediendo es que hay una cuestión técnica de la LAN desde la cual se dispara el *traceroute* que dificulta alcanzar el *gateway*, pues 80ms parece un tiempo elevado para esto. Independientemente, se probó llegar al mismo destino desde otra LAN con mejor tiempo de llegada al *gateway*, y sin embargo también se obtuvo un outlier en un hop posterior, reafirmando el punto anterior de que lo que se considera un outlier depende fuertemente de la escala de la ruta.

Universidad de Moscú

Recorrido en el Planisferio

A continuación se puede ver de forma bastante clara como el paquete tuvo que pasar por estados unidos para luego ir a Europa Occidental hasta llegar finalmente a Rusia.

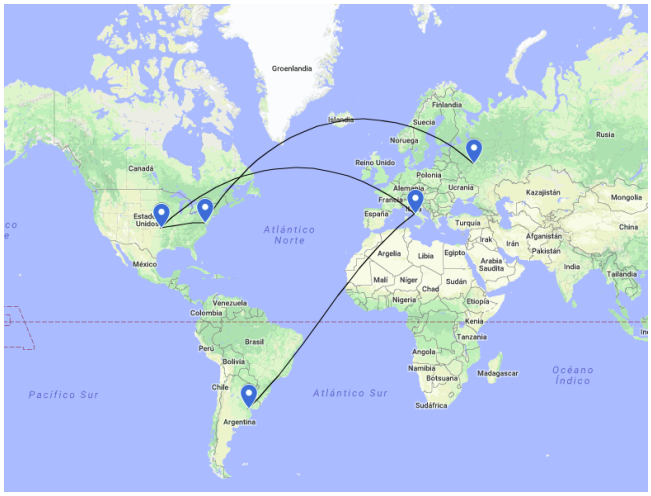

Figura 3: Recorrido realizado por los paquetes durante la ejecución de traceroute al intentar alcanzar el sitio `www.msu.com`

RTT entre saltos

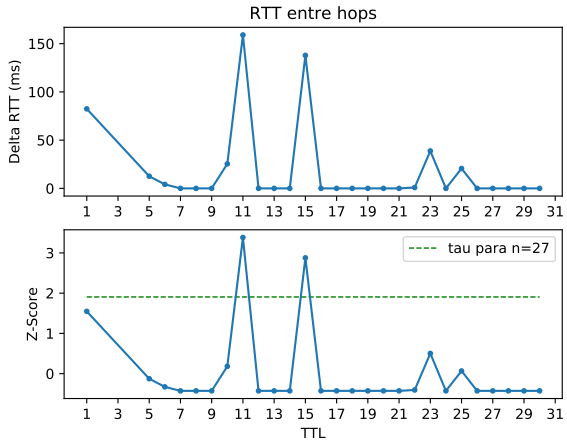

Figura 4: RTT entre saltos (antes y después de normalizar respectivamente) para el sitio `www.msu.com`. n es la cantidad de TTLs para los que se obtuvo un *time exceed* (que son los puntos que se grafican).

Para este caso, podemos observar en la figura 4, que se detectan dos saltos intercontinentales. Este caso tiene bastantes particularidades, en parte a la ruta compleja que se vió en el punto anterior.

El primer pico, que se da en el salto 11, coincide con un salto a un router de Roma, Italia (`89.221.41.171`). La cuestión es que el nodo desde el cual se realiza el hop también está ubicado en Italia (`185.70.203.32`), y dicho salto no fue detectado como outlier, pues en efecto tiene un RTT mucho más bajo. Una posible hipótesis que manejamos al respecto de porque ocurrió esto (que no se detecte el salto continental, pero sí el siguiente hop) es que el primer router al que se llega sea uno de los principales gateways de Italia, lo que provoque que se encuentre congestionado y el paquete que se manda quede encolado un largo tiempo antes de forwardearse al siguiente hop.

El segundo pico, con TTL 15, presenta una situación similar: el destino está en Estados Unidos, pero los tres saltos anteriores también, y no fueron detectados. Acá hay una diferencia sin embargo: esos tres saltos mostraron todos RTTs diferenciales de 0, lo que implica que de hecho el RTT promedio de estos nodos fue menor que el promedio del último router que estaba en Italia. Este parece ser un caso de çamino asimétrico"(Jobst 2012). Razonablemente existe una mejor ruta desde Estados Unidos a Argentina, que cruzar toda Europa. Que la ruta no haya ido directamente por Estados Unidos, salteandose Italia, suena a una consecuencia de un cambio del estado de la red: por alguna razón el camino "óptimo"desde Italia a Rusia dejo de estar habilitado.

En definitiva, hay que decir que en este caso tuvimos dos falsos positivos y tres falsos negativos (Argentina-Italia, Italia-USA, USA-Rusia).

Algo para señanalar, es que este caso se corrió desde la misma LAN que la ruta anterior, y puede verse en los gráficos que efectivamente el RTT del primer salto fue igual, pero en este caso, a diferencia del anterior, no se consideró un *outlier*, pues había saltos mucho más costosos.

Universidad de Sidney

Recorrido en el Planisferio

A continuación se puede ver de forma bastante clara como el paquete tuvo que pasar por estados unidos para luego ir a Europa Occidental hasta llegar finalmente a Rusia.

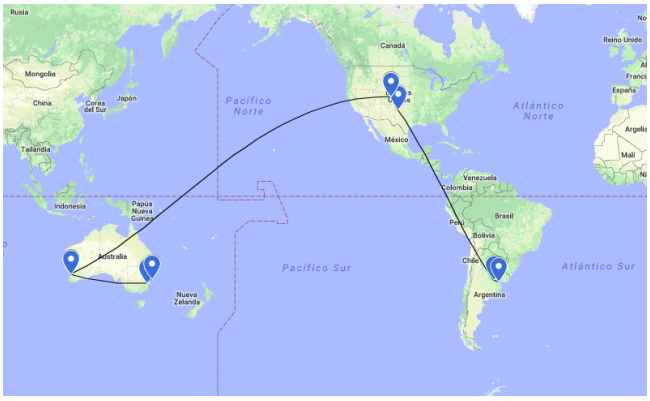

Figura 5: Recorrido realizado por los paquetes durante la ejecución de traceroute al intentar alcanzar el sitio `sydney.edu.au`

RTT entre saltos

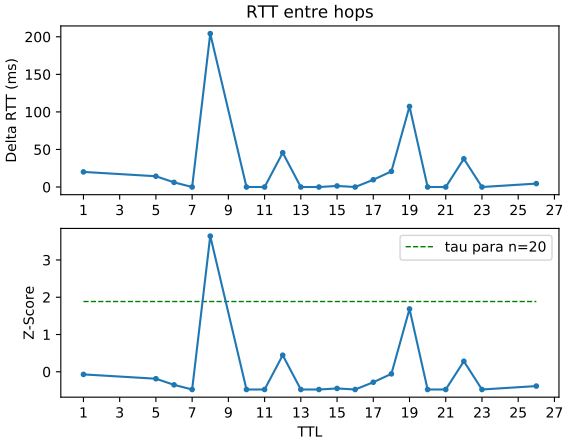

Figura 6: RTT entre saltos (antes y después de normalizar respectivamente) para el sitio `sydney.edu.au`. n es la cantidad de TTLs para los que se obtuvo un *time exceed* (que son los puntos que se grafican).

Finalmente, para esta ruta vemos que se detectó un salto intercontinental en el octavo hop. Este efectivamente

fue un salto de Argentina a Estados Unidos (que como mencionamos en la sección anterior, vamos a considerar un salto continental). Por otro lado, no se detectó el salto de Estados Unidos a Australia en el salto 19, aunque puede verse en el gráfico que estuvo bastante cerca de alcanzar el threshold.

Para esta ruta tuvimos entonces un falso negativo, y un verdadero positivo.

Algo que puede valer la pena mencionar de este caso es que la mayoría de los valles donde el delta RTT da 0ms son saltos donde la diferencia de tiempo es menor que la varianza, lo cual es razonable ya que se mueven entre nodos que están cercanos entre si.

IV. Discusión

Recapitulando, de lo dicho hasta aquí nos llevamos las siguientes conlusiones:

- Vistos los casos de Brasil y Rusia, podemos decir que la calidad de la detección automática depende en altísima medida de la longitud total de la ruta. Se podría concluir que un método estadístico de estas características puede preferirse para casos en los que se sabe de antemano que deben existir enlaces intercontinentales y se desea precisar cuáles son. Esto se debe a que la probabilidad de tener falsos positivos es bastante alta en caso de que no los haya, como una consecuencia de la arbitrariedad de las topologías de redes, y la existencia de posibles congestiones.
- También vimos, con el caso de Rusia, que este modelo es muy sensible a anomalías como los caminos asimétricos (que sospechamos que fue uno de los problemas), o las congestiones en la red (otra suposición).
- Aún en el caso de Australia, donde no parecieron verse anomalías grandes, el modelo tuvo problemas para inferir uno de los saltos continentales.

En definitiva, notamos que este modelo tiene muchas oportunidades de mejora, y no es lo suficientemente robusto como para poder ser usado seriamente como un detector de enlaces intercontinentales. De hecho, salvo que se cuente con cierta metadata sobre el estado de la red, parece imposible tener un buen predictor, considerando la gran variedad de situaciones que se dan actualmente en las redes y que pueden meter ruido, desde congestiones hasta anomalías debidas a las topologías o los protocolos heterogeneos.