Teoría de las Comunicaciones: TP2

Federico Sabatini, Pedro Boitier, Dylan Socolobsky, Ignacio Losiggio

Introducción:

En este trabajo analizaremos diversas rutas que pueden llegar a utilizar los paquetes al circular de un punto del planeta a otro. Tomando como referencia distintas universidades distribuidas a lo largo del mundo realizaremos un análisis sobre las distintas trazas que realizan 2 dispositivos al comunicarse.

Para ello realizaremos nuestros propios scripts que implementan `traceroute`, con el cual obtendremos información de los saltos que se realizan hasta llegar al destino. Se eligieron como candidatos a los experimentos las siguientes universidades:

1. **Argentina**: unc.edu.ar (Universidad de Córdoba)

2. **Islandia:** fs.ru.is (University of Reykjavik)
3. **Egipto:** alexu.edu.eg (University of Alexandria)
4. **Rusia:** itmo.ru (University of Saint Petersburg)
5. **Sudáfrica:** unisa.ac.za (University of South Africa)

En un principio decidimos analizar únicamente universidades que se encuentren en continentes distintos al nuestro, buscando poder observar comportamientos anómalos. Pero decidimos que no vendría mal incluir un caso trivial sin enlaces interocéanicos, como lo es en este caso la Universidad de Córdoba.

Implementación del código

El código original provisto por la cátedra envía paquetes proponiendo un TTL incremental esperando en cada caso la respuesta con el mensaje *Time Exceeded,* para finalmente guardar los valores de TTL y RTT correspondientes.

Decidimos que el TTL se incremente 30 veces, es decir empezando con $TTL = 1$ hasta $TTL = 30$, en cada caso enviando una ráfaga de 30 paquetes. Es decir envíamos un total de $30 * 30 = 900$ paquetes por cada host elegido.

Se calculan 2 RTT diferentes para cada TTL enviado.

1. **RTT** average for the most common answer (el promedio de la respuesta más común)
2. **RTT** for all answers (el promedio de todas las respuestas recibidas)

Los resultados se ordenan dado un dominio de menor a mayor por su respectivo TTL para obtener la variación del RTT entre un salto a otro utilizamos el TTL inmediato anterior (pueden haber **missing hops**).

Por último, fijadas las *IPs* de cada TTL usamos la db de **GeoLite2**[1] para geolocalizar las rutas que recorren nuestros paquetes.

Características de los experimentos

La captura del tráfico fue realizada el *2021-05-22* entre las *19:28hs* y las *22:20hs* desde una red en CABA provista por Fibertel.

Se capturaron tanto los paquetes enviados cómo los recibidos. Los resultados se encuentran en `scripts/traceroute.pickle`. A la hora de calcular los RTTs se descartan todos los mensajes de tipo **Echo-Reply**. Además, a la hora de calcular la cantidad de Hops faltantes se descartan en nuestra experimentación todos los TTL mayores o iguales al primer TTL del que sólo recibimos **Echo-Reply.**

Buscaremos realizar un diagnóstico general de dichos dominios, detectar posibles anomalías siguiendo el paper de *Traceroute Anomalies*[2], así mismo buscaremos identificar los saltos interoceánicos (si es que los tiene) en la red.

---

[1] ref: https://dev.maxmind.com/geoip/geoip2/geolite2/
[2] ref: https://www.net.in.tum.de/fileadmin/TUM/NET/NET-2012-08-1/NET-2012-08-1_02.pdf

Experimentación

Acercamiento a la universidad de Córdoba

Nos gustaría hacer un primer acercamiento eligiendo un único dominio. En este caso elegimos a la Universidad Nacional de Córdoba (unc.edu.ar). Suponemos que no deberían aparecer enlaces interoceánicos ni esperar graves anomalías; esperamos que la respuesta sea un “caso feliz”:

TTL	**RTT** promedio	**RTT** entre saltos	**#Time-Exceeded**	**#Hosts**
1	44.10	8.29	30	1
6	55.35 / 52.40[3]	0.24	30	2
7	53.36 / 52.64	5.69	30	2
8	58.33	7.24	30	1
9	65.57	-	30	1
10	65.31	-	30	1
12	Se recibe un **Echo-Reply**			

Ya nuestro primer acercamiento nos hace notar ciertos comportamientos. El primero y más notable es quizás que en este caso las rutas encontradas sólo difieren en los hops seis y siete, finalizando siempre con los mismos saltos.

Otra cosa interesante es el hecho de que tenemos **missing hops**, dado que no registramos ninguna respuesta para los TTLs entre 2 y 5. Perdiendo toda la información en ese rango. Este suceso suele ocurrir porque un firewall protege esa parte de la red o el enrutamiento utiliza MPLS para esa sección. En este caso vemos la segunda opción más lógica para las circunstancias dadas.

Finalmente, el dominio unc.edu.ar tiene tres IPs cómo registro primario en DNS[4]. Nuestro script envió **Echo-Request** a estas tres direcciones distintas pero siempre recibió respuestas la misma “última milla” de la ruta. Lo que nos lleva a pensar que la/las computadoras que sirven a esas IPs están geográficamente cerca una de la otra.

---

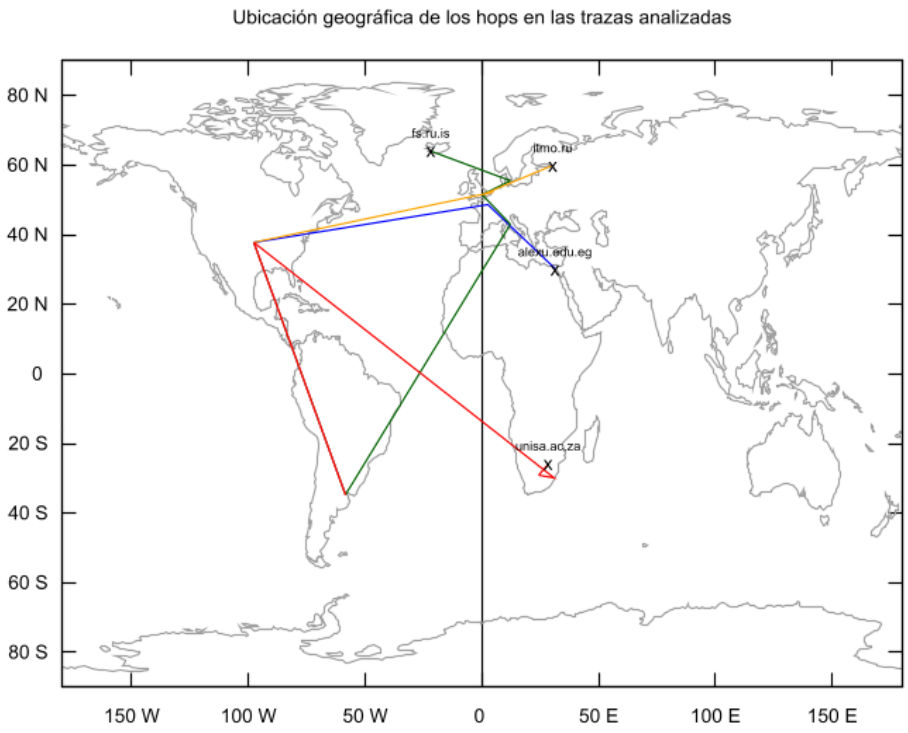
[3] Promedio de todos los hosts / Promedio del host que más veces respondió

[4]
```
$ dig unc.edu.ar
[...]
;; ANSWER SECTION:
unc.edu.ar.             2116    IN      A       200.16.16.171
unc.edu.ar.             2116    IN      A       200.16.16.170
unc.edu.ar.             2116    IN      A       200.16.16.174
[...]
```

Análisis Intercontinental.

A continuación decidimos extender nuestro análisis a universidades más lejanas, pues concluimos que la universidad de Córdoba aunque útil para un análisis elemental se encuentra demasiado cerca y sus respuestas son demasiado triviales.

Para ello usaremos las cuatro universidades restantes mencionadas en la introducción.



Si observamos la ubicación geográfica de las rutas encontradas, podemos notar que se deberían poder detectar saltos interocéanicos. Presentamos sin más los RTT entre saltos para las universidades mencionadas.

TTL	*unisa.ac.za*[5]	*alexu.edu.eg*	*itmo.ru*	*fs.ru.is*[6]
1	244.32	498.86	297.88	94.82
6	12.06 [14]	-[7]	-	138.29 [8]
7	110.63	40.86	130.17	219.17

[5] Nunca se recibió un **Echo-Reply**, pero creemos que el **Echo-Request** se enrutó correctamente pese a eso dado que hay un servidor https en ese dominio que es accesible.
[6] Misma acotación.
[7] Los campos marcados con "-" son aquellos en los que el RTT promedio para este TTL era mayor a todos los RTT promedio posteriores, desestimamos los valores negativos.

8		0.68 [10]	43.10 [10]	1.87
9	4.39	2.41	45.19	25.95
10	86.73 [14]	22.46	-	14.14 [15]
11	123.08 [14]	8.11	-	1.35
12	0.34	68.70		0.61 [14]
13	149.16	12.82	**Echo-Reply**	1.50
14	23.98	35.11		18.54
15	-	1.36		-
16	-	7.74		
17	4.49	35.65		
18		2.87		
24				
25		**Echo-Reply**		

En amarillo figuran aquellos tiempos entre saltos que se calcularon contra un salto que no era el siguiente del que se disponían datos. En verde se marcan los TTL para los que se recibió un *Time-Exceeded* por última vez en la traza. Dado que el enunciado solicita sólo trabajar con paquetes que sean *Time-Exceeded* no se realizaron los cálculos contra los *Echo-Reply* recibidos. En rojo los TTLs que no son el inmediato siguiente al que esta directamente arriba.

Conclusiones

¿Cuál es la cantidad de saltos que no responden los Time-Exceeded?
¿Cuál es el largo de la ruta en términos de los saltos que sí responden?

Host	Hops con alguna respuesta	Largo en saltos respondidos
unc.edu.ar	54,54%	6
itmo.ru	66,66%	8
fs.ru.is	75%	12
unisa.ac.za	72,22%	13
alexu.edu.eg	62,5%	15

Una de las primeras observaciones que notamos al realizar este experimento es la alta cantidad de saltos no respondidos que hemos tenido en nuestros dominios.

Además una de las primeras suposiciones que teníamos al iniciar la experimentación es que la cantidad de saltos totales que debe realizar una red podría ser significativa en la cantidad de mensajes no respondidos de la red (a mayor distancia, más mensajes perdidos)

Sin embargo no llegamos a observar dicha correlación, la realidad es que la mayoría de los saltos no respondidos ocurren en los primeros 6 hops antes de siquiera salir de Argentina. Se puede observar este comportamiento viendo que la universidad de Córdoba en efecto más de la mitad de sus saltos no han respondido.

¿La ruta tiene enlaces intercontinentales? ¿Cuántos?

Si, todos los dominios exceptuando la UNC poseen enlaces intercontinentales, ascendiendo a un total de 8 enlaces distintos.

- **itmo.ru:** posee 1 o 2 enlaces intercontinentales:
    - Argentina a Estados Unidos seguido de Estados Unidos a Amsterdam
    - Argentina a Amsterdam

  De Amsterdam es redirigido finalmente a Rusia. La razón de esta duda yace en la diferencia mostrada entre la geolocalización de la IP y el RDNS de la misma: la geolocalización dice que la IP pertenece a un lugar de Estados Unidos mientras que el RDNS dice ser parte de la infraestructura de Level3 en Amsterdam. Creemos que la información de geolocalización es una medida más confiable en este caso.

  Éste mismo problema ocurrirá varias veces durante la sección. La razón principal de esto es que la información de Geolocalización viene muchas veces de la asignación original de las IPs por ICANN, entonces una empresa Estadounidense puede muchas veces terminar usando IPs de Estados Unidos en su infraestructura en terreno Europeo. O (por ejemplo) Telecom Italia usando IPs Italianas en la llegada del Atlantis-2 en Conil, España.

- **fs.ru.is:** Lo logra en 1 enlace intercontinental, primero avanza desde Argentina a Italia (en realidad entendemos que a España por las razones antes mencionadas) y de ahí redirige (aparentemente) por Reino Unido, Dinamarca y finalmente Islandia.

- **unisa.ac.za:** Posee 2 enlaces, avanza desde Argentina a Estados Unidos y luego desde Estados Unidos a Sudáfrica (Durban según geoip, aunque Mtunzini es más probable) desde dónde se dirige al destino.

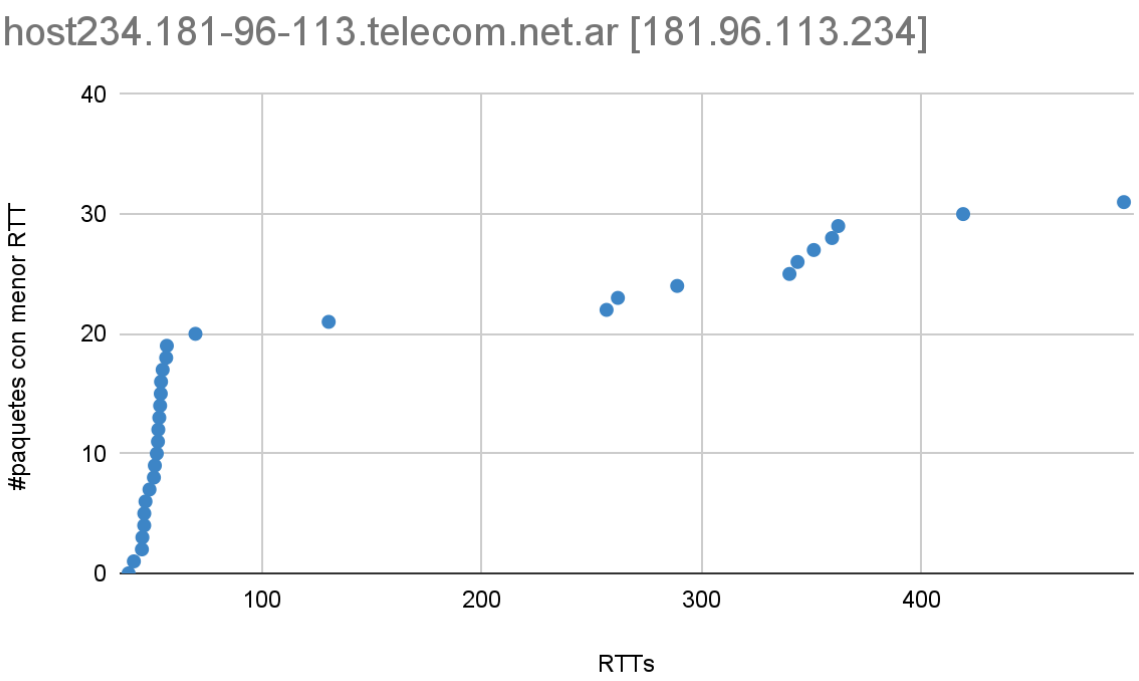
- **alexu.edu.eg:** Posee 3 enlaces, Argentina a Estados Unidos, luego desde Estados Unidos llega a un host de telecom en Francia con el cual finalmente redirecciona a Egipto. Aquí otra vez si nos dejáramos llevar por los RDNS de los hops podríamos sumar la duda de si el hop en francia (telecom-egypt.demarc.cogentco.com [149.14.135.106]) está realmente en Francia, creemos que sí dados todos los cables en el mediterráneo que van casi directamente de Marsellas, Francia a Alejandría, Egipto.

¿Se observaron comportamientos anómalos del tipo descrito en la bibliografía sugerida [6]?

- **Missing hops:** Estuvieron presentes en todas las capturas realizadas.
- **Missing destination:** Tanto en *fs.ru.is* cómo *unisa.ac.za* no respondían a los *Echo-Request*. Pese a eso entendemos que los hosts están online y los paquetes fueron enrutados correctamente, dado que podemos acceder a las páginas web hosteadas en las IPs correspondientes.
- **False Round-Trip Times:** Los más obvios son al llegar al sexto hop, el cuál siempre fué la IP `181.96.113.254 (host254.181-96-113.telecom.net.ar)` para todas las universidades en otros países.
  Éste host sólo respondió 32 de los 120 paquetes que se enviaron con ese TTL hacia universidades en el exterior. De éstos 21 tienen menos de 70ms de RTT pero los 11 restantes tienen entre 130ms y 500ms. Creemos que hay alguna asimetría en los caminos de retorno de este host que hace que muchas veces éstos se pierdan y que incluso cuando llegan ⅓ de las veces lo hacen mucho más tarde de lo esperable.
- **False links:** No hicimos un análisis suficientemente exhaustivo para determinar cuáles links son posiblemente falsos, pero en la experimentación que realizamos no observamos TTLs con respuestas tanto del tipo *Time-Exceeded* cómo *Echo-Reply*, de ésto podemos cómo mucho concluir que en caso de estar mezclando distintas rutas en nuestras trazas al menos éstas tienen la misma longitud.
  Cómo el fin de este trabajo no era descubrir la topología de la red en cuestión no ahondamos más en esto.

¿Se observaron otros comportamientos anómalos? Proponga hipótesis que permitan explicarlos.

Tras nuestras observaciones hemos encontrado que existen routers con comportamientos aparentemente bimodales, Por ejemplo el router de telecom que responde al $TTL = 6$ el RTT promedio da un valor considerablemente alto. Nuestras hipótesis fueron propuestas en la sección anterior.


host234.181-96-113.telecom.net.ar [181.96.113.234]
40
30
20
10
0
#paquetes con menor RTT
100
200
300
400
RTTs

Detección de enlaces interoceánicos

¿La distribución de RTT entre saltos presenta outliers según el método de Cimbala? ¿Cuántos?

Para responder esta pregunta volvamos a la tabla de RTT-entre saltos antes presentada:

TTL	*unisa.ac.za*	*alexu.edu.eg*	*itmo.ru*	*fs.ru.is*
1	244.32	498.86	297.88	94.82
6	12.06 [14]	-	-	138.29 [8]
7	110.63	40.86	130.17	219.17
8		0.68 [10]	43.10 [10]	1.87
9	4.39	2.41	45.19	25.95
10	86.73 [14]	22.46	-	14.14 [15]
11	123.08 [14]	8.11	-	1.35
12	0.34	68.70		0.61 [14]
13	149.16	12.82	**Echo-Reply**	1.50
14	23.98	35.11		18.54
15	-	1.36		-
16	-	7.74		
17	4.49	35.65		
18		2.87		
24				
25		**Echo-Reply**		

Ahora la gama de colores violetas representa si la celda dada fué el primer, segundo o tercer outlier encontrado. Dada la forma del algoritmo propuesto por la cátedra podemos intuir que cuánto *“antes”* se descubre un outlier más *“foráneo”* es respecto del resto de los elementos.

¿Se corresponden los outliers con los enlaces intercontinentales? ¿Cuántos falsos positivos y falsos negativos hay?

Sólo tres: el doceavo de *alexu.edu.eg*, el sexto[8] de *fs.ru.is* y el séptimo de *fs.ru.is*, los otros tres resultan falsos negativos.

¿Se aprecia alguna diferencia en la capacidad de detectar enlaces intercontinentales según el largo de la ruta?

No. Se puede notar que no se encuentran outliers para la ruta de *itmo.ru*, suponemos que debido a que sólo cuatro de los RTTs entre hops se pudieron calcular (el resto de los TTLs tenían RTT promedio mayor a todos los hops posteriores).

¿Es posible mejorar las predicciones usando un valor de corte fijo para el valor $\frac{X_i - \bar{X}}{S}$ en lugar del valor en la tabla τ?

No realmente. Intentamos encontrar valores que expusieran los enlaces que creíamos interoceánicos haciendo búsqueda binaria sobre τ pero todo valor de corte fijo terminaba agregando muchos más saltos que no creemos tiene sentido que sean interoceánicos.

[8] El sexto salto cuenta cómo intercontinental porque el RTT entre saltos calculado se hizo contra el rtt promedio del TTL=8, el cual creemos está en otro continente.

Anexo: trazas recolectadas

Dadas las peculiaridades de la información recolectada creemos importante agregar la información que tuvimos en mente para la confección de este informe.

unisa.ac.za

TTL	Host, IP, Ciudad y País	RTT Promedio
1	_gateway [192.168.1.1]	82.24
6	host234.181-96-113.telecom.net.ar [181.96.113.234], Argentina	326.56
7	[8.243.138.29], Buenos Aires, Argentina	136.87
9	[4.68.37.222], United States	247.50
10	ae-3.r23.miamfl02.us.bb.gin.ntt.net [129.250.3.141], United States	251.89
11	ae-2.a01.miamfl02.us.bb.gin.ntt.net [129.250.3.167], United States	215.54
12	xe-0-0-23-2.a01.miamfl02.us.ce.gin.ntt.net [129.250.200.114], United States	189.11
13	ae0-42.rt04.bb.ampath.net [190.103.184.13], United States	189.46
14	187-185-103-190.ampath.net [190.103.185.187], United States	338.63
15	et-0-1-0-0-isd1-pe1.net.tenet.ac.za [155.232.1.149], Durban, South Africa	364.61
16	et-0-0-0-0-pta2-pe3.net.tenet.ac.za [155.232.1.97], Durban, South Africa	362.44
17	et-0-0-2-0-pta2-pe2.net.tenet.ac.za [155.232.128.79], South Africa	355.41
18	te0-6-0-2-pta2-pe1.net.tenet.ac.za [155.232.128.91], South Africa	359.90

alexu.edu.eg

TTL	Host, IP, Ciudad y País	RTT Promedio
1	_gateway [192.168.1.1]	40.10
6	host234.181-96-113.telecom.net.ar [181.96.113.234], Argentina	538.96
7	[8.243.138.29], Buenos Aires, Argentina	143.99
8	ae-2-52.ear4.Miami2.Level3.net [4.69.207.41], United States	184.85
9	be3356.ccr22.mia03.atlas.cogentco.com [154.54.10.57], United States	183.12
10	be2025.ccr21.mia01.atlas.cogentco.com [154.54.47.229], United States	185.54
11	be3482.ccr41.atl01.atlas.cogentco.com [154.54.24.145], United States	208.00
12	be2112.ccr41.dca01.atlas.cogentco.com [154.54.7.157], United States	216.11
13	be2331.ccr31.bio02.atlas.cogentco.com [154.54.85.242], United States	284.82
14	be2324.ccr31.mad05.atlas.cogentco.com [154.54.61.130], United States	297.64
15	be3355.ccr31.vlc02.atlas.cogentco.com [154.54.57.230], United States	295.74
16	be3222.ccr21.mrs01.atlas.cogentco.com [154.54.57.205], United States	289.36
17	be2189.rcr21.mil01.atlas.cogentco.com [154.54.36.70], United States	297.11
18	telecom-egypt.demarc.cogentco.com [149.14.135.106], France	332.76
24	[193.227.16.32], Egypt	335.64

itmo.ru

TTL	Host, IP, Ciudad y País	RTT Promedio
1	_gateway [192.168.1.1]	39.94
6	host234.181-96-113.telecom.net.ar [181.96.113.234], Argentina	337.82
7	[8.243.138.29], Buenos Aires, Argentina	140.25
8	ae-2-3204.edge7.Amsterdam1.Level3.net [4.69.162.181], United States	270.42
9	BR2.Amsterdam1.surf.net [213.244.164.2], Amsterdam, Netherlands	268.33
10	ae2-11.RT.OV.SPB.RU.retn.net [87.245.232.0], United Kingdom	313.52
11	GW-ITMO.retn.net [87.245.250.79], United Kingdom	312.63
12	[77.234.192.167], St Petersburg, Russia	310.10

fs.ru.is

TTL	Host, IP, Ciudad y País	RTT Promedio
1	_gateway [192.168.1.1]	38.55
6	host234.181-96-113.telecom.net.ar [181.96.113.234], Argentina	133.38
7	[195.22.220.56], Italy	52.51
8	[195.22.209.198], Italy	271.68
9	[195.66.225.24], United Kingdom	273.56
10	is-rey.nordu.net [109.105.102.180], Denmark	299.52
11	ndn-gw1.rhnet.is [109.105.102.181], Denmark	293.15
12	Neshagi01-te1-11.rhnet.is [130.208.17.106], Reykjavik, Iceland	294.50
13	Taeknigardur00-te1-11.rhnet.is [130.208.17.70], Reykjavik, Iceland	293.61
14	Hringbraut00-te0-2.rhnet.is [130.208.17.134], Reykjavik, Iceland	295.12
15	RHnet-gw2.ru.is [130.208.18.38], Reykjavik, Iceland	313.66
16	[130.208.247.4], Reykjavik, Iceland	311.63