DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 2

Rutas en Internet

8 de diciembre de 2015 — Teoría de las comunicaciones

Integrante	LU	Correo electrónico
González, Sergio Martín	723/10	sergiogonza90@gmail.com
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 del siguiente trabajo, es el de estudiar y poder monitorear rutas a nivel de red, y por sobre todo, encontrar rutas hacia hosts en otros continentes, por las que solo se puede acceder a través de enlaces submarinos. La idea es intentar encontrar estos enlaces, e intentar monitorear como se comportan a lo largo del dia.

Para esto, es necesario implementar una herramienta *traceroute* basada en el protocolo ICMP (Internet Control Message Protocol), y que esta se pueda utilizar para la recopilación de los datos necesarios para luego poder hacer los análisis pertinentes. Este protocolo, el cual es el utilizado en la herramienta *ping* de cualquier sistema operativo, es un protocolo de control y notificación de errores, que corre sobre IP. ICMP cuenta con varios tipos de paquetes, pero nosotros solo vamos a enfocarnos en 3:

- ***echo-request***: Paquete utilizado en la herramienta *ping*. Este sirve para saber si un host se encuentra disponible o no. De estar disponible, el host receptor responde con un *echo-reply*.
- ***echo-reply***: Respuesta al envio del tipo de paquete anterior. Cuando un host recibe un paquete de tipo *echo-request*, este envia (de tenerlo habilitado) un paquete de este tipo.
- ***time-exceeded***: Los paquetes ICMP poseen un campo llamado TTL (time to live), el cual indica el tiempo de vida del paquete. Entonces, por ejemplo, si se envia un *echo-request* con un TTL de 3, luego de los 3 saltos de router, este paquete es descartado, y se envía un paquete de tipo *time-exceeded* al nodo que envió el paquete original.

Utilizando esta ultima propiedad de los paquetes ICMP, podemos implementar una herramienta que, envíe paquetes *echo-request* incrementando de a poco el TTL (inicialmente con 1), y quedarnos con las IPs origen de los paquetes *time-exceeded*, asi poder averiguar la IP de cada salto que realiza el paquete al momento de ser enviado. Esto se realiza hasta que se obtiene un paquete de tipo *echo-reply*.

También se deben poder obtener los ΔRTT entre cada hop de la ruta, para poder realizar un test de Grubbs, y encontrar posibles outliers, los cuales son potenciales enlaces submarinos. Para realizar esto, lo que se hace es acumular los RTT de cada respuesta de tipo *time-exceeded*, y luego:

$$\Delta RTT_i = RTT_i - RTT_{i-1}$$

En la etapa de análisis, una ves ya teniendo las la información necesaria, se hara un análisis sobre la posible ubicación de los nodos que se encuentran entrelazados por un enlace submarino. Para esto se utilizara la herramienta Geoiptool [1] con la que se puede geolocalizar una IP.

Luego se realizar un análisis sobre la variación del RTT del enlace submarino, a lo largo del día. Para esto, se realizara un script que corra el traceroute cada 30 minutos, y asi poder realizar un monitoreo sobre la ruta.

[1] www.geoiptool.com/es/

2. Primera Consigna: Caracterizando rutas

2.1. Implementacion de Traceroute

Para el desarrollo de los próximos análsis, se implemento una herramienta *traceroute* sobre Scapy en Python 3. La idea consiste en lo siguiente: Enviar iterativamente paquetes ICMP *echo-request*, empezando con un TTL de 1 y incrementándolo hasta que nos llegue un paquete de tipo *echo-reply*. Por cada paquete enviado, nos guardamos la IP del host originario del paquete, y de esta manera podemos construir la ruta que lleva al mismo.

Sin embargo, hay cosas a tener en cuenta para poder realizar esta técnica. Primero, es posible que por el trafico de la red en el momento de hacer la captura, o por otros motivos que están fuera de nuestro control, el paquete pueda cambiar de ruta en los distintos instantes del envio de los paquetes ICMP. Otra cosa a tener en cuenta, es que si nosotros queremos hacer un estudio sobre el RTT de cada enlace (ΔRTT), no nos sirve el RTT medido para 1 solo intento, ya que esto pudo también haber sido influenciado por el trafico momentáneo de la red.

Para intentar aliviaran esto, lo que hacemos es, no enviar solo 1 paquete por cada TTL, sino que enviamos varios. Cuantos?, los necesarios para poder realizar un promedio lo mas confiable posible del ΔRTT. Lo que hacemos es, establecer una cota inferior y una cota superior a la cantidad de paquetes a enviar de un mismo TTL:

- **MAX_ATTEMPTS**: Cantidad máxima de paquetes **exitosos** a enviar de un mismo TTL. Esto quiere decir que, si mandamos MAX_ATTEMPTS paquetes, y siempre nos llega una respuesta del mismo nodo, entonces promediamos el RRT, anotamos esta IP, y pasamos al próximo TTL.
- **MIN_ATTEMPTS**: Cantidad mínima de paquetes provenientes del mismo nodo que debemos recibir para pasar al próximo TTL. Esto quiere decir, que si nosotros enviamos MIN_ATTEMPTS paquetes, y todos vinieron del mismo nodo, y en MIN_ATTEMPTS + 1, nos viene de un nodo diferente a los anteriores, de todas maneras promediamos los RTT, y avanzamos de TTL. En caso de que la IP cambie antes de llegar a los MIN_ATTEMPTS, se descarta la información recopilada para este TTL y se vuelve a comenzar.

Cabe aclarar que este método igual tiene una falla, y es que si al momento de cambiar la IP, nosotros igual proseguimos al próximo TTL, o mismo cuando no llegamos a los MIN_ATTEMPTS, y debemos reiniciar el TTL, va a haber un ΔRTT que quizá no sea el de 1 enlace físico, ya que la ruta cambio. Por eso es que junto con la información recopilada, nos quedamos también con los intentos que se realizaron para cada salto, así podemos identificar cuando pase esto. Como el objetivo del TP en si es estudiar enlaces submarinos, no es tan importante que la ruta cambie en un momento determinado, siempre y cuando esto no suceda en los extremos del enlace submarino.

2.2. Identificando enlaces Submarinos

Para poder identificar enlaces submarinos, nos basamos en un teste denominado "Test de Grubbs", el cual sirve para detectar outliers en una muestra aleatoria. En nuestro caso, la muestra aleatoria seran los ΔRTT obtenidos en todo el traceroute, y el outlier que queremos encontrar, sera justamente el del enlace submarino, ya que estimamos que este enlace debe tener un RTT mucho mayor a cualquier enlace que se encuentre en tierra.

Para realizar el test de Grubbs, es necesario que la muestra aleatoria sea de distribucion normal, pero de todas maneras, la cátedra indico que igual usemos el test si la distribucion de los ΔRTT no nos queda normal. Ahora para verificar si la distribución de nuestras muestras es normal, utilizaremos el modulo Scipy de Python, el cual permite hacer análisis estadísticos. El test de Grubbs consiste en lo siguiente, dada una muestra X de tamaño N:

1. Se calcula el promedio muestral de X: μ
2. Se calcula la desviacion standard de X: σ
3. Se calcula el estadistico del test: $G = (max(X) - \mu)/\sigma$
4. Se calcula el valor de rechazo del test: : $C = \frac{n-1}{\sqrt{N}} * \sqrt{\frac{t^2_{\alpha/N,N-2}}{N-2+t^2_{\alpha/N,N-2}}}$
5. Si el valor de G es mayor a C, entonces *max(X)* es un outlier de X

El valor $t_{\alpha/N,N-2}$ hace referencia al valor critico de la distribución t para N-2 grados de libertad, y nivel significativo α

2.3. Rutas a explorar

Teniendo implementadas las herramientas para recopilar la información y realizar los cálculos anteriormente dichos anteriormente, debemos elegir 2 IPs ubicadas en otro continente para poder realizar el *traceroute*. Las IPs elegidas son las siguientes:

- Finlandia: University of Helsinki - `www.helsinki.fi`
- Inglaterra: University of Oxford - `www.ox.ac.uk`

Una vez corrido el traceroute a estas IPs, tendremos una ruta aproximada hacia las mismas. Cabe aclarar que la herramienta no es exacta, sino que solo es una aproximación de la ruta real, ya que para poder encontrar la ruta exacta con todos los nodos involucrados, es necesario usar técnicas mas complejas.

Una ves tengamos encontrada una ruta, lo que hacernos es geolocalizar las IPs de la misma, para poder comparar esto con el resultado que nos devuelva el Test de Grubbs, y asi corroborar que RTT del enlace que vemos como outlier, tenga sus extremos en distintos continentes. Para esto, utilizaremos Geo IP Tool (`www.geoiptool.com`), junto con IP Location (`www.iplocation.net`), las cuales nos permiten ubicar geográficamente una IP. IP Location, a diferencia de Geo IP Tool, presenta 4 posibles opciones de donde puede estar la IP, sacadas de otras paginas Web.

3. Segunda Consigna: Gráficos y Análisis

Luego de la breve explicacion de como se han implementado las herramientas a utilizar, y de haber propuesto 2 rutas a analizar, procedemos a mostrar los resultados obtenidos de las mismas. Para poder realizar los siguientes analisis, lo que hicimos fue: Correr la herramienta *traceroute* a lo largo del dia, cada media hora con un script, durante 12 horas.

3.1. Rutas encontradas

Primero, vamos a presentar las rutas encontradas para las IPs propuestas. Se presentaran en formato tabla, y se indicara la ubicacion de cada IP segun Geo IP Tool.

3.1.1. Universidad de Helsinki (Finlandia)

El siguiente cuadro muestra como el enlace submarino se debería encontrar entre Argentina y Italia. Esta ruta, fue el resultado de 27 corridas de *traceroute* a lo largo del dia, por lo que es lo mas certero que da nuestra herramienta. En la ruta se ve como al llegar al continente europeo, las IPs van saltando entre IPs locales de Suecia y Finlandia, hasta llegar al host destino.

Una cosa a destacar, es la aparicion de varios ΔRTT negativos, incluso habiendo promediado los RTT de 10 paquetes distintos, y que ademas como se puede ver, todos fueron exitosos, y solo hubo 1 cambio de IP al inicio del *traceroute*, en donde luego de 8 intentos cambio la IP. Esto muestra como por mas de que se promedien varios RTT, la cercanía de estos nodos, hacen que los RTT varíen mucho. Para eso esta bueno tambien tener el valor del desvió standard de los intentos, asi tenemos una aproximación de entre que valores puede estar verdaderamente el RTT, con alta probabilidad.

Cuadro 1: Ruta para Universidad de Helsinki

TTL	IP	Intentos	RTT Promedio	Desvio Standard	Delta RTT	Ubicación
1	10.0.0.1	8	59.12ms	15.03ms	59.12ms	Router Local (Argentina)
2	10.24.128.1	10	55.40ms	4.30ms	-3.73ms	Router Local (Argentina)
3	181.47.254.85	10	58.20ms	6.09ms	2.80ms	Argentina (Bs As)
4	195.22.220.93	10	60.10ms	6.01ms	1.90ms	Argentina (Tigre)
5	195.22.220.92	10	63.11ms	7.41ms	3.01ms	Argentina (Tigre)
6	**149.3.183.11**	**10**	**271.50ms**	**4.70ms**	**208.39ms**	**Italia**
7	No hubo respuesta	-	-	-	-	-
8	109.105.97.126	10	324.10ms	2.60ms	52.60ms	Suecia
9	109.105.102.102	10	336.80ms	8.99ms	12.70ms	Suecia
10	109.105.102.103	10	333.20ms	17.01ms	-3.60ms	Suecia
11	193.167.253.9	10	333.80ms	8.51ms	0.60ms	Finlandia
12	128.214.173.242	10	336.40ms	3.44ms	2.60ms	Finlandia (Helsinki)
13	128.214.173.10	10	333.22ms	12.11ms	-3.18ms	Finlandia (Helsinki)
14	128.214.189.85	10	328.40ms	5.15ms	-4.82ms	Finlandia (Helsinki)
15	128.214.189.90	10	337.20ms	5.81ms	8.80ms	Finlandia (Helsinki)

3.1.2. Universidad de Oxford (Inglaterra)

A diferencia de la la ruta vista en el punto anterior, esta resulto tener algunas particularidades. Para empezar, el enlace que tiene el mayor ΔRTT, tiene ambos extremos en Estados Unidos, uno en Virginia y el otro en Kansas (segun Geo IP Tool y IP Location). Por otro lado, hay otro salto grande que se observa, que es entre el hop 7 y 9, y aquí si ambos extremos estan en distintos continentes. Nunca conseguimos que el hop 8 responda el paquete ping, ya que muy probablemente sea un router que descarta los paquetes de tipo ICMP. Esto es un comportamiento bastante habitual que tiene algunos routers.

De todas maneras, según se ve, el enlace debe estar entre Estados Unidos e Inglaterra. Probando tambien para otras IPs aledaneas a la utilizada, obtuvimos los mismo resultados, por lo que optamos por igual quedarnos con esta.

Cuadro 2: Ruta para Universidad de Oxford

TTL	IP	Intentos	RTT Promedio	Desvio Standard	Delta RTT	Ubicación
1	10.0.0.1	8	55.38ms	10.89ms	55.38ms	Router Local (Argentina)
2	10.24.128.1	10	60.00ms	3.77ms	4.62ms	Router Local (Argentina)
3	181.47.254.85	10	58.50ms	5.78ms	-1.50ms	Argentina (Bs As)
4	208.178.195.214	10	65.30ms	5.85ms	6.80ms	Estados Unidos (Virginia)
5	208.178.195.213	10	65.10ms	6.72ms	-0.20ms	Estados Unidos (Virginia)
6	67.17.75.66	10	182.60ms	14.40ms	117.50ms	Estados Unidos (Kansas)
7	4.68.111.121	10	188.30ms	5.23ms	5.70ms	Estados Unidos (Chicago)
8	No hubo respuesta	-	-	-	-	-
9	**212.187.139.166**	**10**	**273.50ms**	**7.68ms**	**85.20ms**	**Inglaterra (Londres)**
10	146.97.33.2	10	273.10ms	6.92ms	-0.40ms	Inglaterra (Londres)
11	146.97.37.194	10	276.70ms	7.89ms	3.60ms	Inglaterra (Londres)
12	193.63.108.94	10	273.40ms	4.48ms	-3.30ms	Inglaterra (Gales)
13	193.63.108.98	10	273.90ms	5.88ms	0.50ms	Inglaterra (Gales)
14	193.63.109.42	10	271.20ms	5.51ms	-2.70ms	Inglaterra (Gales)
15	192.76.21.71	10	273.70ms	5.31ms	2.50ms	Inglaterra (Oxford)
16	192.76.22.200	10	275.90ms	5.15ms	2.20ms	Inglaterra (Oxford)
17	192.76.32.62	10	283.50ms	37.29ms	7.60ms	Inglaterra (Oxford)
18	129.67.242.154	10	270.90ms	4.20ms	-12.60ms	Inglaterra (Oxford)

3.2. Test de Grubbs

En los resultados que mostramos anteriormente, se vio el ΔRTT del enlace submarino para una sola corrida del *traceroute*, pero esto fue solo para dar un primer vistaso y una primera estimacion sobre donde podria encontrarse el enlace submarino. Lo que ahora queremos mostrar, es el resultado que dio el Test de Grubbs, sobre los enlaces anteriormente encontrados, y ver si efectivamente son outliers, y potenciales enlaces submarinos.

ACLARACION: Si bien las tablas no es un buen metodo para presentar datos, en esta ocasión decidimos si utilizarla, ya que se pueden ver todos los valores para todas las corridas del test de Grubbs, esto son, el p-valor del *normaltest*, el estidistico G y el valor para la condicion de rechazo C.

Como se puede ver, el test de grubbs indica que los ΔRTT del enlace visto en el punto anterior (desde 195.22.220.92 a 149.3.183.11), son outliers de la muestra. Esto reafirma el hecho de que aqui se encuentre un enlace submarino. Tambien se ve como en 1 solo caso, el test indico que no hay outliers en la muestra, y ademas, el ΔRTT mas grande, no es el enlace visto anteriormente.

Si se realiza una mirada mas de cerca sobre los resultados de esta ejecución, se podrá ver como tiene varios ΔRTT muy grandes en comparación a otras ejecuciones. De echo esto hace que el *normaltest* indique que es una distribución normal, pero sin embargo, el test de grubbs indica que el máximo valor no es un outlier. Debido a estas inconsistencias con respecto a las demás ejecuciones, se opto por ignorar esta ejecución, y asumir que fue influenciada por el estado de la red particular de ese momento.

3.2.1. Universidad de Helsinki (Finlandia)

Cuadro 3: Grubbs para Universidad de Finlandia

HORA	IP Extremo 1	IP Extremo 2	DRTT	p-valor	G	C	Es outlier?
19:57	128.214.173.10	128.214.189.85	333.70	0.588223	1.54	15.895	NO
20:30	195.22.220.92	149.3.183.11	212.10	0.000000	3.25	15.895	SI
20:30	195.22.220.92	149.3.183.11	209.70	0.000000	3.25	15.895	SI
21:01	195.22.220.92	149.3.183.11	216.70	0.000000	3.26	15.895	SI
21:33	195.22.220.92	149.3.183.11	214.30	0.000001	3.22	15.895	SI
22:05	195.22.220.92	149.3.183.11	207.10	0.000001	3.20	15.895	SI
22:36	195.22.220.92	149.3.183.11	207.60	0.000001	3.23	15.895	SI
23:08	195.22.220.92	149.3.183.11	210.00	0.000001	3.21	15.895	SI
23:39	195.22.220.92	149.3.183.11	207.70	0.000001	3.23	15.895	SI
00:11	195.22.220.92	149.3.183.11	210.60	0.000000	3.25	15.895	SI
00:42	195.22.220.92	149.3.183.11	213.30	0.000000	3.24	15.895	SI
01:13	195.22.220.92	149.3.183.11	208.39	0.000000	3.25	15.895	SI
01:45	195.22.220.92	149.3.183.11	210.00	0.000001	3.24	15.895	SI
02:48	195.22.220.92	149.3.183.11	206.20	0.000001	3.22	15.895	SI
03:19	195.22.220.92	149.3.183.11	210.70	0.000001	3.22	15.895	SI
03:51	195.22.220.92	149.3.183.11	206.30	0.000000	3.25	15.895	SI
04:22	195.22.220.92	149.3.183.11	206.30	0.000001	3.21	15.895	SI
04:54	195.22.220.92	149.3.183.11	209.70	0.000000	3.25	15.895	SI
05:25	195.22.220.92	149.3.183.11	209.70	0.000001	3.23	15.895	SI
05:57	195.22.220.92	149.3.183.11	211.00	0.000000	3.24	15.895	SI
06:28	195.22.220.92	149.3.183.11	205.90	0.000000	3.24	15.895	SI
07:00	195.22.220.92	149.3.183.11	205.30	0.000000	3.23	15.895	SI
07:31	195.22.220.92	149.3.183.11	214.00	0.000000	3.24	15.895	SI
08:02	195.22.220.92	149.3.183.11	227.70	0.000000	3.25	15.895	SI
08:34	195.22.220.92	149.3.183.11	205.70	0.000001	3.23	15.895	SI
09:05	195.22.220.92	149.3.183.11	209.00	0.000001	3.23	15.895	SI
09:37	195.22.220.92	149.3.183.11	217.70	0.000002	3.17	15.895	SI

3.2.2. Universidad de Oxford (Inglaterra)

En este caso, el test de Grubbs dio que el enlace con mayor ΔRTT era un outlier, pero como vimos con las herramientas de geolicalizacion de IPs, este enlace probablemente no sea un enlace submarino. Debido a esto, optamos por volver a ejecutar el test de Grubbs, pero esta vez quitando este outlier, para tratar de identificar la presencia de un segundo outlier, y asi si ver si segun el test, el enlace que nosotros creemos que es el enlace submarino, es tambien un outlier segun el test de Grubbs. Los resultados pueden verse en la tabla 4

Cuadro 4: Grubbs para Universidad de Finlandia

HORA	IP Extremo 1	IP Extremo 2	DRTT	p-valor	G	C	Es outlier?
20:31	4.68.111.121	212.187.139.166	86.90	0.100648	2.48	15.971	SI
21:02	4.68.111.121	212.187.139.166	91.40	0.000021	3.02	15.971	SI
21:34	4.68.111.121	212.187.139.166	80.90	0.000311	2.85	15.971	SI
22:05	4.68.111.121	212.187.139.166	73.80	0.001433	2.69	15.971	SI
22:37	4.68.111.121	212.187.139.166	85.00	0.000052	2.95	15.971	SI
23:09	4.68.111.121	212.187.139.166	92.70	0.000007	3.12	15.971	SI
23:40	4.68.111.121	212.187.139.166	85.10	0.000090	2.92	15.971	SI
00:43	4.68.111.121	212.187.139.166	86.00	0.000011	3.01	15.971	SI
01:14	4.68.111.121	212.187.139.166	83.10	0.000012	3.02	15.971	SI
01:46	4.68.111.121	212.187.139.166	88.20	0.000008	3.08	15.971	SI
02:17	4.68.111.121	212.187.139.166	82.30	0.000008	3.03	15.971	SI
02:49	4.68.111.121	212.187.139.166	80.30	0.000110	2.89	15.971	SI
03:20	4.68.111.121	212.187.139.166	81.40	0.000015	2.99	15.971	SI
03:52	4.68.111.121	212.187.139.166	81.57	0.000021	3.00	15.971	SI
04:23	4.68.111.121	212.187.139.166	88.70	0.000120	2.92	15.971	SI
04:55	4.68.111.121	212.187.139.166	82.30	0.000048	2.96	15.971	SI
05:26	4.68.111.121	212.187.139.166	85.20	0.000011	3.05	15.971	SI
05:58	4.68.111.121	212.187.139.166	84.70	0.000014	3.00	15.971	SI
06:29	4.68.111.121	212.187.139.166	84.30	0.000019	2.99	15.971	SI
08:03	4.68.111.121	212.187.139.166	108.10	0.000178	2.97	15.971	SI
08:34	193.63.108.98	193.63.109.42	141.50	0.042960	2.28	15.971	SI
09:06	4.68.111.121	212.187.139.166	87.30	0.000010	3.04	15.971	SI
09:38	4.68.111.121	212.187.139.166	85.30	0.000018	2.98	15.971	SI

4. Conclusiones