Topological and Geographical Analysis on Routing and Server Selection of Anycast Clouds

Felipe Espinoza
NIC Chile Research Lab
Santiago, Chile
fdns@niclabs.cl

Abstract

Anycast is used by a large amount of DNS and CDN services for distribution, load balancing and latency reduction in the access of its resources. Given the private nature of the internet, the knowledge about the area of services of every anycast server and the optimality of the routes used by their clients to connect to these servers become very diffuse. This paper presents initial work on a geographical and topological mapping of an anycast cloud over the entire Internet, dividing it into /24 networks prefixes, presenting methods to perform measurements of routes, area of service and distances between servers and their clients, comparing the theorical optimum and the observed reality.

1 Introducción

Anycast es una metodología de enrutamiento que permite la distribución de los paquetes a diferentes servidores utilizando una única dirección o bloque IP. Esta metodología es utilizada por servidores *DNS* (*Domain Name System*, por sus siglas en inglés) [ICCCN6] y *CDN* (*Content Delivery Network*, por sus siglas en inglés) [CONEXT15] para realizar una distribución de la carga que estos reciben, y acercar los servicios a los clientes, permitiendo disminuir los tiempos de acceso a los diferentes recursos, e incrementar su disponibilidad ante problemas que puedan ocurrir en la red.

Para permitir a los clientes conectarse a los servidores más cercanos, cada *AS* (*Autonomous System*, por sus siglas en inglés) utiliza el protocolo *BGP* (*Border Gateway Protocol*, por sus siglas en inglés) para definir el camino más corto de un paquete para llegar a su destino. Sin embargo, estos sistemas no presentan formas de asegurar que las rutas seleccionadas por los *AS* abarquen a la cantidad óptima de usuarios en la cual operan.

Dado el carácter privado de las características de la red y las tablas de rutas de los *AS*, realizar una predicción del camino y servidor al cual cada cliente se conectará pasa a ser una tarea muy difícil. Esta información es muy importante para los proveedores de estos servicios, para la detección de errores, balanceo de carga, y posicionamiento de nuevos servidores en puntos estratégicos que permitan el mejor rendimiento y latencia para los clientes.

NIC Chile actualmente opera su propia nube anycast con más de 26 servidores distribuidos en diferentes países, lo cual incrementa la complejidad del ruteo y su descubrimiento. Para obtener un mayor conocimiento sobre las rutas actuales, este trabajo busca realizar una exploración sobre las diferentes rutas seleccionadas por los distintos *AS* en una dirección de red anycast, determinando el área de servicio de cada servidor, los *AS* involucrados y características de la red sobre la que se trabaja.

2 Trabajo Relacionado

Los análisis que se han realizado sobre las redes anycast utilizan distintos métodos para obtener el área de servicio. Dos de los métodos más utilizados en esta área corresponden a los siguientes.

- **Análisis de logs y capturas de paquetes**: Esta forma de análisis permite a los operadores

In: Proceedings of the IV School of Systems and Networks (SSN 2018), Valdivia, Chile, October 29-31, 2018. Published at http://ceur-ws.org

de los servicios anycast obtener las fuentes desde donde se genera el tráfico de manera directa. Esta posee la ventaja de entregar información exacta de los lugares en los que sus servicios son utilizados, y los volúmenes de información sobre cada una de las redes [PAM07]. Este tipo de análisis no entrega información sobre la topología de la red sobre la que se trabaja, lo cual no nos permite realizar un análisis a fondo sobre las rutas que se utilizan.

- **Redes de medición**: Utilizando plataformas con una gran cantidad de puntos de medición, es posible realizar un análisis del estado de una red anycast de manera externa. Este tipo de análisis realiza mediciones de valores tales como el *RTT* (*Round Trip Time*, por sus siglas en inglés) de diferentes localizaciones, o utiliza propiedades del software ejecutado en la nube anycast, tal como el envío de consultas *CHAOS* a servidores *DNS* para la identificación de estos.

  Este tipo de mediciones permite obtener una visión externa de los servicios. Sin embargo, no logran realizar una observación completa sobre los clientes de la nube anycast, dado que los puntos de observación no existen en todas las redes de internet. Plataformas tales como *RIPE Atlas* proveen aproximadamente 10,000 puntos de observación que permiten una buena visión sobre el despliegue [PAM17], sin embargo, generalmente estos se encuentran sesgados a lugares con una gran concentración de usuarios, principalmente Estados Unidos y Europa, lo cual produce que mediciones en lugares tales como Sudamérica y Asia sean poco representativas.

Por otro lado, se han realizado evaluaciones a largo plazo sobre la disponibilidad y efectividad de los diferentes despliegues de redes anycast, mostrando los efectos de diferentes eventos sobre cambios en la red [NOMS10]. Además, se han realizado estudios sobre la estabilidad de estas redes, donde se ha encontrado que las rutas anycast son mayoritariamente estables, permitiéndonos realizar una mejor planificación de los despliegues de estos servidores [TNSM17], sin esperar cambios importantes en la topología que afecten al rendimiento de esta.

Las últimas investigaciones utilizan redes de medición, realizando una comparación entre el caso óptimo y el observado. En caso de poseer acceso a la red anycast, se han desarrollado mediciones utilizando paquetes *ICMP* para establecer los bloques IP y áreas de servicio de cada servidor anycast [IMC17.2].

3 Metodología Propuesta

En este trabajo, se busca responder las siguientes preguntas: (a) ¿Son los servidores más cercanos los que responden las consultas *DNS*? (b) ¿Es posible determinar las rutas utilizadas por los clientes para conectarse a los servidores DNS? (c) ¿Es posible determinar el posicionamiento de un nuevo servidor DNS según la topología de internet?

3.1 Medición de rutas y área de servicio

El análisis de las áreas de servicio de cada servidor se separará en prefijos de red /24, dado que este corresponde al límite actual utilizado para el establecimiento de las rutas *BGP*. En cada uno de estos prefijos, se elegirá a un representante que se encuentre activo para realizar las mediciones. Para esto, es posible utilizar *Hitlists* ya calculadas o realizar un cálculo local de esta [IMC10].

Para ejecutar las mediciones del área de servicio de cada dirección IP. Un servidor en la red anycast a analizar realizará el envío de un paquete *UDP* o *ICMP* a cada dirección de la *Hitlists*, estableciendo en los demás servidores sondas que reciban las respuestas de los paquetes enviados. Dado que la dirección de respuesta corresponde a la nube anycast, los paquetes serán dirigidos por la red hacia el servidor correspondiente al cliente en el prefijo de red indicado, tal como se puede apreciar en la figura 1.

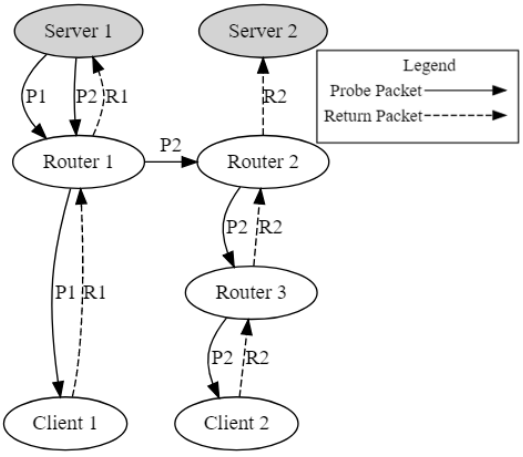

Figura 1: Ejemplo de ejecución de medición de área de servicio. El servidor 1 envía paquetes a los cliente 1 y 2, y las respuestas son enrutadas según las tablas de rutas de los routers a los servidores respectivos.

Para realizar el cálculo de la ruta sobre la cual cada servidor en la nube anycast trabaja, por cada paquete recibido del análisis, es necesario ejecutar un *traceroute* para identificar los *routers* intermedios desde cada punto de medición. Este proceso puede optimizarse

realizando una predicción sobre la cantidad de saltos que separan el cliente y el servidor observando el *TTL* (*Time To Live*, por sus siglas en inglés) recibido desde el cliente, tomando la suposición de que los clientes generalmente utilizan un *TTL* de 64, 128 o 255. Con este *TTL*, es posible ejecutar el *traceroute* desde el cliente hasta el origen, y detener su ejecución al encontrar un router común con otras mediciones, permitiendo evitar el cálculo de rutas ya realizadas, como se aprecia en la Figura 2.

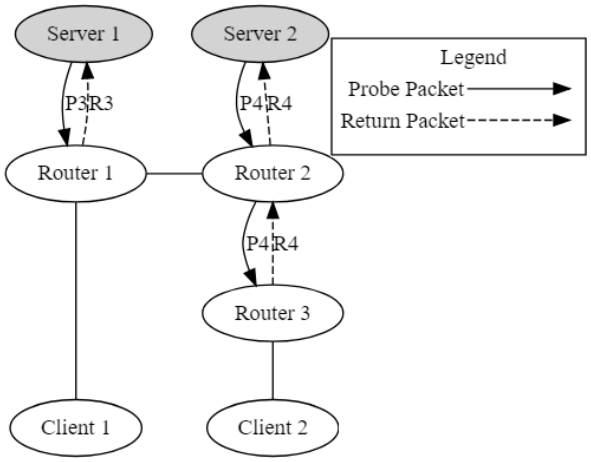

Figura 2: Segundo trazado ejecutado por los servidores receptores, el cual se realiza desde un paso anterior del *TTL* recibido, hasta detectar un nodo ya medido. Si el *Router 3* ya se encuentra medido, no se realizara mediciones al *Router 2* u otros anteriores.

3.2 Medición de distancias geográficas y topológicas

Para generar un mapa topológico sobre las distintas redes sobre las cuales cada servidor en la nube anycast trabaja, es posible asociar cada dirección IP en las rutas calculadas a un *AS* en particular utilizando información recolectada por *RIPE NCC* sobre los *Servicios de Ruteo de Información* (*RIS*, por sus siglas en inglés, *Routing Information Service*), la cual posee las subredes sobre las cuales cada *AS* trabaja.

Para establecer las distancias actuales entre cada paso de las rutas calculadas, se utilizará el *RTT* obtenido de cada medición, estableciendo la distancia entre cada *router* o *AS* como el tiempo que toma cada paquete de pasar desde un salto a otro.

Para realizar la medición de las distancias geográficas del servidor y sus clientes, se utilizará la base de datos *MaxMind* [MAXM] para realizar la geolocalización a nivel país de cada dirección IP analizada. En caso de necesitar incrementar la precisión a nivel ciudad, es posible utilizar versiones pagadas de *MaxMind* o *NetAcuity* [NETAC], las cuales ofrecen una precisión de 70.1% y 66.5% respectivamente [IMC17.1].

3.3 Calculo de selección de servidor óptimo

Para realizar la comparación de las rutas teóricas y calculadas desde el punto de vista topológico, se utilizarán las rutas extraídas desde el análisis, y se complementarán con bases de datos externas tales como tablas de ruta *RIS* de *RIPE NCC* y la base de datos *CAIDA's Ark* [CAIDA]. Luego de esto, se utilizarán algoritmos de búsqueda para calcular los pares óptimos entre clientes y servidores, tomando en cuenta las distancias calculadas entre los *AS*.

Para el caso de la comparación geográfica, se utilizará la localización de los servidores anycast ya conocida, y se comparará con la localización de los clientes obtenida utilizando la base de datos *MaxMind* a nivel de ciudad. Este resultado se comparará con las distancias a los otros servidores, determinando si el calculado corresponde o no al más cercano.

4 Resultados Futuros

A través de los experimentos a realizar, esperamos obtener información sobre las diferentes áreas de servicio de cada servidor, determinando los routers o *AS* en los cuales se produce el cambio de rutas de un servidor anycast a otro, además de todos los routers intermedios que se encuentran asociados a cada servidor.

Por otro lado, podremos apreciar el porcentaje de clientes que poseen las rutas óptimas a los servidores anycast, en términos de distancia topológica y geográfica, apreciando las diferencias que pueden existir entre estas.

Utilizando la información sobre las distancias topológicas, en complementación con información de los clientes actuales en los servidores, será posible calcular una posición en la red tal que, al posicionar un nuevo servidor en la nube anycast, la reducción de los *RTT* sea máxima. Esta posición solo podrá calcularse de forma teórica, dado a diferencias en los algoritmos utilizados para el cálculo de las rutas en los diferentes *AS*. Sin embargo, nos entregara información inicial para comenzar un análisis de la posición de un nuevo servidor.

Por último, es posible utilizar la información generada por estos experimentos para realizar otros tipos de análisis, tal como la detección de *IP Spoofing*, realizando una comparación entre la IP de los clientes con el área de servicio de cada servidor, logrando detectar consultas que no deberían ser recibidas por estos. Esto puede ser utilizado como medida de mitigación de ataques *DDoS*, donde servidores *DNS* son utilizados como método de amplificación.

Referencias

[ICCCN6] S. Sarat, V. Pappas and A. Terzis, *"On the Use of Anycast in DNS"*. Proceedings of 15th International Conference on Computer Communications and Networks, Arlington, VA, 2006, pp. 71-78.

[CONEXT15] Danilo Cicalese, Jordan Augé, Diana Joumblatt, Timur Friedman, and Dario Rossi. *"Characterizing IPv4 anycast adoption and deployment"*. In Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT '15). ACM, New York, NY, USA, 2015, Article 16, 13 pages.

[IMC10] Fan, X., and Heidemann, J. *Selecting representative IP addresses for Internet topology studies*. In Proceedings of the ACM Internet Measurement Conference (Melbourne, Australia, Nov. 2010), ACM, pp. 411–423.

[MAXM] MaxMind Inc. 2018. MaxMind GeoIP2 City. https://www.maxmind.com/en/geoip2-databases. (Agosto 2018).

[NETAC] Digital Envoy. 2018. Digital Element NetAcuity databases. https://www.digitalelement.com/geolocation/. (Agosto 2018).

[IMC17.1] Manaf Gharaibeh, Anant Shah, Bradley Huffaker, Han Zhang, Roya Ensafi, and Christos Papadopoulos. 2017. *A look at router geolocation in public and commercial databases*. In Proceedings of the 2017 Internet Measurement Conference (IMC '17). ACM, New York, NY, USA, 463-469.

[CAIDA] The CAIDA Internet Topology Data Kit - 2017-08, http://www.caida.org/data/internet-topology-data-kit

[NOMS10] Bu-Sung Lee, Yu Shyang Tan, Y. Sekiya, A. Narishige and S. Date, *"Availability and effectiveness of root DNS servers: A long term study"*, 2010 IEEE Network Operations and Management Symposium - NOMS 2010, Osaka, 2010, pp. 862-865.

[TNSM17] L. Wei and J. Heidemann, *"Does anycast hang up on you?,"* 2017 Network Traffic Measurement and Analysis Conference (TMA), Dublin, 2017, pp. 1-9.

[IMC17.2] Wouter B. de Vries, Ricardo de O. Schmidt, Wes Hardaker, John Heidemann, Pieter-Tjerk de Boer, and Aiko Pras. 2017. *Broad and load-aware anycast mapping with verfploeter*. In Proceedings of the 2017 Internet Measurement Conference (IMC '17). ACM, New York, NY, USA, 477-488.

[PAM07] Liu Z., Huffaker B., Fomenkov M., Brownlee N., claffy . (2007) *"Two Days in the Life of the DNS Anycast Root Servers."* In: Uhlig S., Papagiannaki K., Bonaventure O. (eds) Passive and Active Network Measurement. PAM 2007. Lecture Notes in Computer Science, vol 4427. Springer, Berlin, Heidelberg

[PAM17] Oliveira Schmidt R, Heidemann J, Kuipers J.H. *"Anycast Latency: How Many Sites Are Enough?"*. In: Kaafar M., Uhlig S., Amann J. (eds) Passive and Active Measurement. PAM 2017. Lecture Notes in Computer Science, vol 10176.