Conectividad en la región LAC: 2020

Autor: Agustín Formoso
Coordinación/revisión: Guillermo Cicileo y Elisa Peirano
Edición: Maria Gayo, Carolina Badano, Martín Mañana
Proyecto: Fortalecimiento de Infraestructura de Internet Regional
Area: I+D en Infraestructura de Internet

lacnic
www.lacnic.net

Cuando hablamos de conectividad, ¿de qué estamos hablando? Desde que las redes se conectan unas con las otras, los operadores han estado trabajando para que la conectividad entre ellas sea mejor. En este artículo les presentamos un estudio de conectividad donde les explicamos cómo medimos la conectividad de la región LAC, y cómo ha evolucionado durante los últimos años.

Introducción

El término conectividad es utilizado a menudo en la industria de Internet, pero su significado puede variar según el contexto en el que se esté abordando. Puede ser medida según su capacidad de ancho de banda, su número de saltos, o bien, en el caso de este artículo, según su latencia. En este sentido, decimos que dos lugares se encuentran bien conectados cuando la latencia entre ellos es baja, es decir, el tiempo que le lleva a un mensaje ir de origen a destino es bajo.

Desde LACNIC buscamos conocer en mayor detalle las características de la interconexión de redes en América Latina y el Caribe, con el fin de que los operadores cuenten con información que les sirva a sus estrategias de crecimiento. A partir de ahí, es muy útil realizar mediciones de conectividad que abarquen toda la región, incluida toda la región de El Caribe y no sólo la región LAC como región de RIR. Típicamente las mediciones de conectividad se realizan entre un origen y un destino, o entre pocos orígenes y pocos destinos; por lo general, las mediciones se lanzan desde nodos en nuestra propia infraestructura, o hacia nuestra propia infraestructura. Pero para obtener mediciones de conectividad que abarquen toda la región es necesario lanzar mediciones desde redes de terceros, lo cual es un desafío que requiere la colaboración de muchos actores. A continuación se ofrece un listado de plataformas que ofrecen la posibilidad de lanzar mediciones desde redes de terceros, y una breve explicación de sus características:

- RIPE Atlas es un claro ejemplo, donde los usuarios que colaboran deciden hospedar una sonda (puede ser de hardware o de software) y mantenerla conectada a Internet. De esa forma otros usuarios de la plataforma pueden utilizar esa sonda para lanzar mediciones. Si bien la presencia de sondas de RIPE Atlas en LAC ha ido mejorando a lo largo del tiempo, aún no hay presencia suficiente como para realizar estudios que abarquen toda la región.
- CAIDA Ark, o Archipielago, es una plataforma similar a RIPE Atlas, basada en Raspberry Pi generación 2. Lamentablemente, sólo cuenta con 11 sondas en 10 sistemas autónomos de la región LAC.
- M-Lab, o Measurement Lab, es una plataforma que ofrece una serie de herramientas para medir diferentes parámetros de red. Lamentablemente, estas herramientas no ponen foco en la medición de latencia entre redes de terceros, sino en mediciones de

ancho de banda a través de NDT (Network Diagnosis Tools) entre el cliente y los servidores de M-Lab (que no son muchos en la región LAC).

- Por otra parte Speedchecker, una plataforma con similares funcionalidades, ofrece mejor cobertura en la región LAC. A diferencia de RIPE Atlas, esta plataforma cuenta únicamente con sondas de software, lo que hace que el sistema sea más inestable que las demás. Sin embargo, dado que en estudios anteriores habíamos utilizado esta plataforma, decidimos utilizarla nuevamente y así poder comparar los resultados con mayor seguridad.

Metodología y estudios anteriores

LACNIC ha llevado a cabo estudios similares durante años anteriores (pueden ver los estudios para los años 2016 (en inglés) y 2017) A grandes rasgos, estos estudios dan una vista panorámica sobre la región en cada año, y muestran sensibles mejoras en las mediciones de año-sobre-año.

La metodología para realizar las mediciones es la siguiente:

1. Desde todos los países de la región LAC, se programan mediciones cada hora.
2. Estas mediciones son pings ICMP, donde se envían 10 paquetes por cada medición.
3. El destino de los paquetes es una dirección IP aleatoria de un pool de direcciones IP conocidas
    a. Este pool de direcciones IP se compone por todos los servidores Speedtest de la región. Esto permite tener una gran cantidad de puntos de prueba, ubicados en muchísimas redes de la región, con un uptime considerablemente bueno (es altamente probable que la dirección IP sea alcanzable al momento de realizar la medición).
4. Se recolectan los datos y se realiza post-procesamiento
    a. La información de geolocalización se obtiene de Maxmind
    b. La información de routing (Sistema Autónomo, prefixo bajo el que se anuncia) se obtiene de RIPE RIS

Estudio 2020

Para esta edición del estudio, se comenzó una campaña de mediciones desde comienzos de septiembre hasta comienzos de noviembre. En ese período se realizaron 13.000 mediciones desde 26 países, con destino a 25 países. A su vez, esas 13.000 mediciones provinieron desde 332 diferentes Sistemas Autónomos.

Resultados

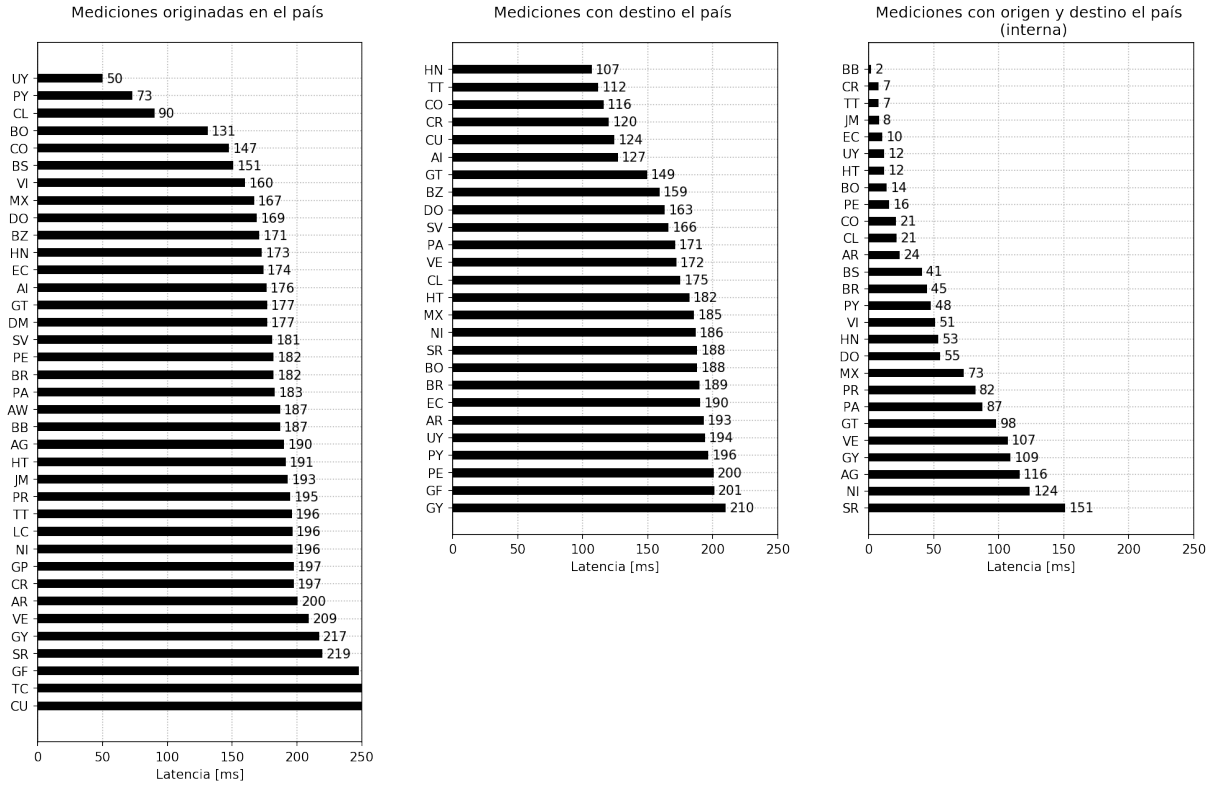
Una vez obtenidos los datos, se agruparon las mediciones en 3 perspectivas diferentes:

1. **País de origen:** esta categoría contiene todas las mediciones con sondas ubicadas en el país, pero no incluye aquellas que terminan en el mismo país. Es decir, las mediciones son únicamente *salientes* del país.

2. **País de destino:** esta categoría contiene todas las mediciones con puntos de prueba (servidores) ubicados en el país, pero no incluye aquellas que son originadas en el mismo país. Es decir, las mediciones son únicamente *entrantes* al país

3. **Ambos países de origen y de destino:** esta categoría contiene todas las mediciones cuyo país de origen y destino es el mismo. Es decir, son mediciones que son *internas* al país. Esta categoría contiene todas las mediciones con sondas y puntos de prueba ubicados dentro del mismo país.

Los resultados se pueden visualizar a través de los siguientes tres gráficos, donde mostramos la mediana de los resultados de cada país:



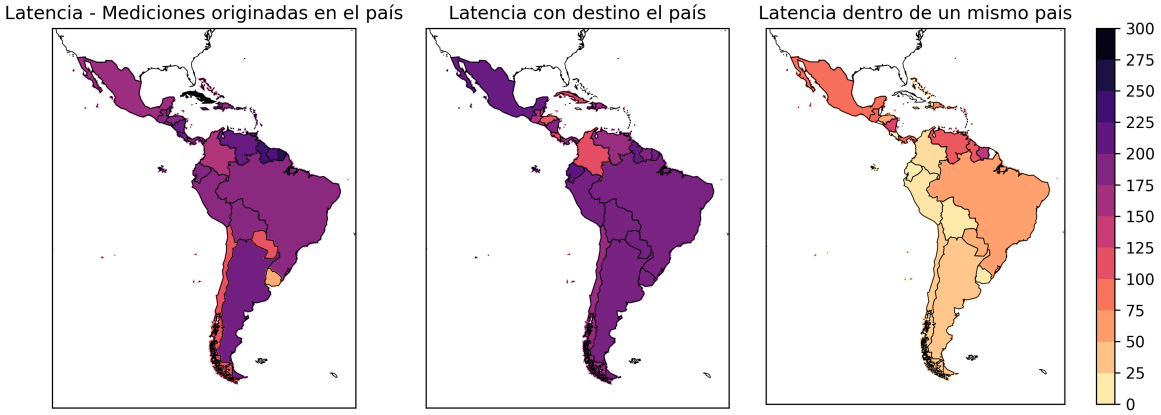
A simple vista se puede ver que tanto para la latencia saliente (1) como para la entrante, los valores rondan entre 150 y 200 milisegundos. Además, naturalmente, se puede ver que la latencia interna (3) es menor a la latencia saliente y la entrante.

Otro aspecto que se puede ver de esta campaña es que la cantidad de países con sondas activas es mucho mayor que la cantidad de países con servidores de pruebas: 37 y 26 respectivamente; esta diferencia hace que el gráfico 1 tenga más entradas que los gráficos 2 y 3. Además, implica que no todas las combinaciones <*país_origen, país_destino*> se encuentren cubiertas por esta campaña de mediciones, al menos hasta el momento.

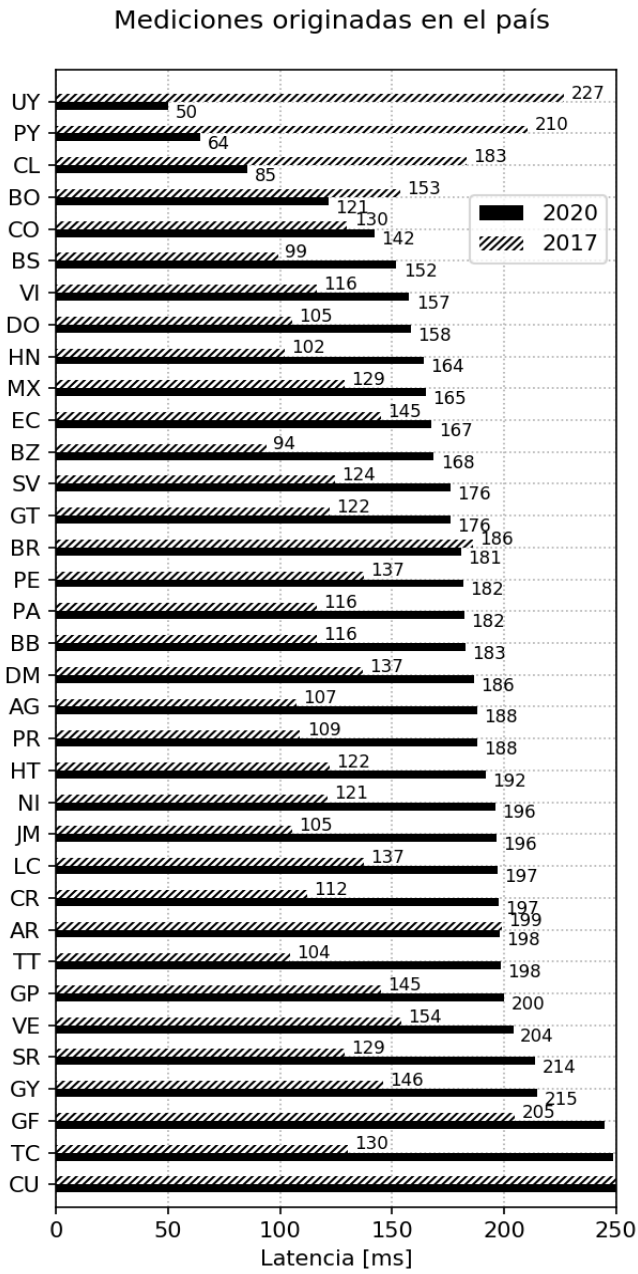
Cabe hacer una mención especial de algunos países que presentan malas latencias en los gráficos anteriores, ya que se debe principalmente a un sesgo en la metodología de mediciones. Como mencionamos anteriormente, el destino de las mediciones es un pool de servidores en la región LAC. Debido a que existe una concentración desigual de esos servidores en la región (hay más servidores en países más grandes), al agendar una nueva medición existe una probabilidad más alta de que sea agendada hacia estos países más grandes. Y de hecho los países con peores latencias del gráfico 1 - Cuba (CU), Islas Turcas y Caicos (TC), Guyana Francesa (GF), Suriname (SR), Guyana (GY), y Venezuela (VE)-, tienen, hasta el momento de este reporte, resultados

únicamente hacia esos países. A medida que pase el tiempo y se agenden más mediciones hacia más países, este sesgo irá bajando y las mediciones reflejarán de mejor forma la realidad de la conectividad regional.

Los gráficos anteriores también se pueden visualizar en un mapa:



Comparando 2020 con 2017



Anteriormente mencionamos que llevamos a cabo un estudio similar durante 2017. ¿Cómo se comparan los resultados obtenidos en 2020 con los anteriores? De similar forma a como se presentaron los resultados en la sección anterior, podemos ver una comparativa de las mediciones agrupadas. Este gráfico representa las mediciones que fueron originadas en cada país.

De la campaña de mediciones de 2017 notamos que hay algunos países que presentan latencias muy por encima de su valor 2020. En particular, los valores para Uruguay (UY), Paraguay (PY), Chile (CL), y Bolivia (BO) del gráfico a la izquierda. Pudimos verificar que no se debe a una baja cantidad de muestras, ya que tienen entre 100 y 1000 muestras. El único país que sí tiene un sesgo de medición es Cuba (CU), que presenta únicamente 1 muestra durante 2017.

Pero observando los datos más detalladamente, vemos que la mayoría de los países presenta en 2020 *peores* valores de latencia hacia la región, en comparación con 2017. De los 37 países del gráfico, 28 (75%) presentan valores más altos. ¿A qué se debe este incremento? Para responder esa pregunta necesitamos hacer un análisis más profundo sobre los datos.

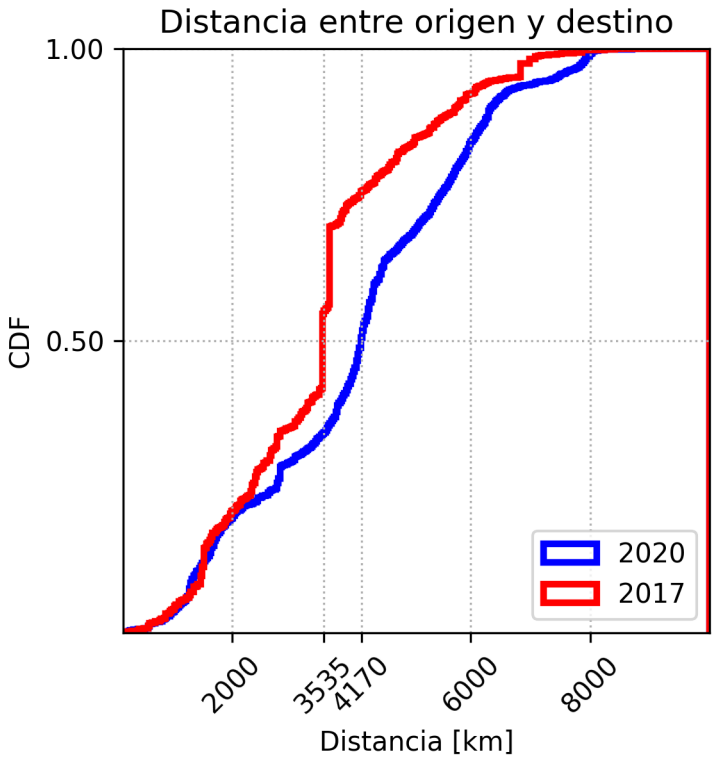
Nos preguntamos: ¿será posible que los países hayan empeorado su conectividad? Nos cuestionamos esa posibilidad, y nos preguntamos: ¿será posible que las mediciones que hemos realizado en 2020 estén recorriendo distancias más grandes que las de 2017? Es una posibilidad, sobre todo si tenemos en cuenta que:

1. Los puntos de prueba de 2020 no son los mismos que los de 2017: ¡se dan de alta y de baja nodos de Speedtest todos los días!
2. Cada medición que agendamos elige de forma *aleatoria* el punto de prueba: existe una componente aleatoria que es fuerte al comienzo del experimento, cuando aún no hemos tenido una cantidad de muestras estadísticamente significante. ¿Podrá ser que de hecho se haya elegido puntos más distantes en el comienzo de este experimento?

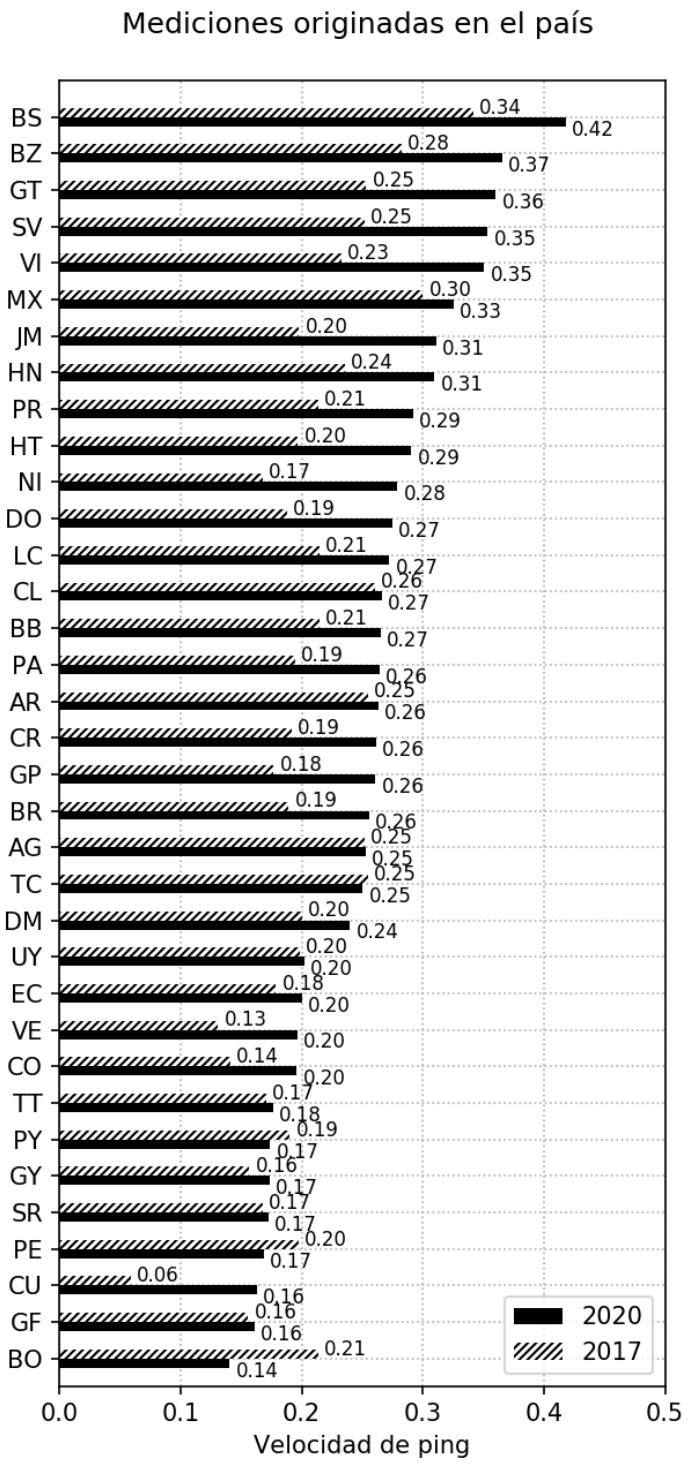
Para responder esa pregunta utilizamos una base de datos de geolocalización[1] para conocer el país de origen y la ciudad de destino de cada medición, y así aproximar, a través de una línea geodésica[2], la distancia que habría recorrido la medición. Si tomamos cada medición, le asignamos una distancia recorrida, y graficamos estas distancias para los años 2017 y 2020, obtenemos una figura como la de abajo, donde las curvas que se encuentran ubicadas más a la izquierda del gráfico representan distancias menores (distancias más cortas). Esta figura es conocida como una CDF[3], y en este caso representa la fracción de mediciones que están por debajo de una determinada distancia. En el eje X tenemos representadas las distancias recorridas y en el eje Y la fracción de distancias que se encuentran por debajo del valor de las X.

El hecho de que la curva roja se encuentre ubicada sobre la izquierda del gráfico nos da un indicio que el set de datos de 2017 contiene distancias más cortas, es decir, puntos que se encuentran más próximos uno de otro, por lo que es lógico que ese set de datos contenga latencias absolutas más bajas. Del gráfico se puede ver que el 50% de las mediciones más cortas de 2017 recorren sólo una distancia de 3535 km, mientras que el 50% de las mediciones más cortas de 2020 recorren una distancia de 4170 km.

[1] Maxmind - versiones junio 2017 y noviembre 2020
[2] Más sobre geodésica en Wikipedia
[3] Más sobre CDF en Wikipedia



Una vez hecha la comparación de distancias, agregamos la componente de latencias. Por ejemplo, para una distancia de 1000 km, no es lo mismo un ping de 10 ms que uno de 20 ms, consideramos que el ping de 10 ms viaja más rápido, es decir tiene mayor velocidad. La forma en que se agrega la componente de latencias es analizando la velocidad de ping, es decir la relación distancia/tiempo de la medición, frente a la velocidad de propagación de la luz en un medio de fibra óptica (una aproximación teórica común es utilizar *c/3*, o un tercio de la velocidad de la luz en el vacío). Esto nos da una fracción, donde un valor cercano a 1 indica que el ping viaja tan rápido como indica la teoría, y valores cercanos a 0 indican velocidades sub-óptimas (debido a caminos más largos que la línea directa, saltos intermedios que introducen retrasos, y otras posibles fuentes de demoras). En este caso, y a diferencia de la medición de latencia, tener un valor más alto es mejor.



Una vez introducida la corrección, calculamos la nueva métrica (velocidad de ping) e hicimos la misma comparación que antes. Podemos ver que ahora 31 de 35 países han percibido mejoras en 2020 frente a 2017 (88%), y únicamente 4 no (Bolivia, Paraguay, Perú, e Islas Turcas y Caicos). Esta nueva métrica nos ayuda a normalizar sets de datos diferentes bajo una misma métrica que los hace comparables, y así poder medir una mejora en la conectividad regional.

Velocidad de ping

Si definimos velocidad como *distancia/tiempo*, podemos decir que un ping que recorre una distancia *d* en un tiempo *t* viaja a una velocidad *d/t*. Usualmente las unidades de distancia son kilómetros (km) y las de tiempo de ping son milisegundos (ms), pero la unidad resultante *km/ms* no tiene mucho sentido. Es por eso que para definir *velocidad de ping* comparamos *d/t* con una velocidad de referencia. Esta velocidad de referencia es *c/3*, la velocidad teórica de propagación de la luz sobre un medio de fibra óptica.

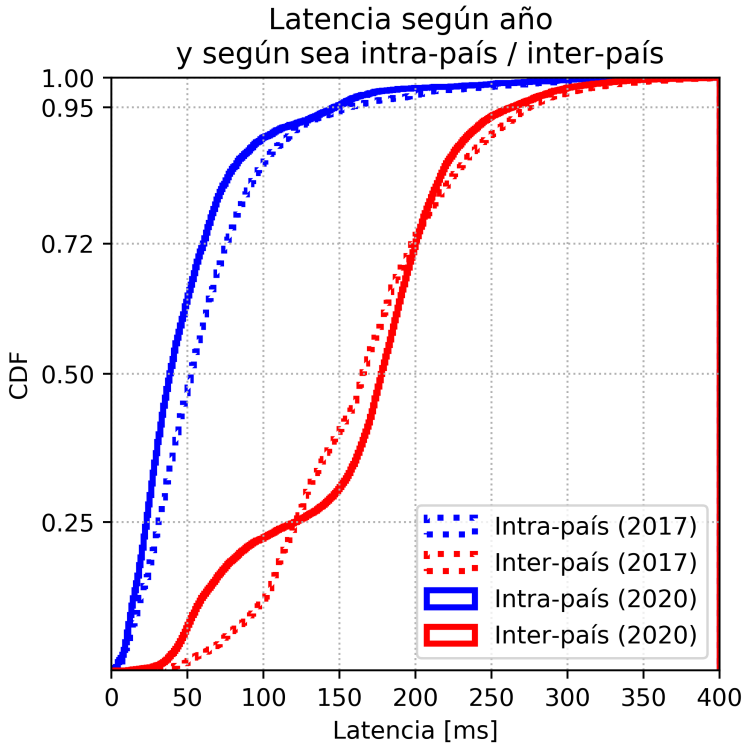
Por lo tanto la comparación resulta en:

$$\frac{d/t}{c/3}$$

Donde la unidad resultante es una fracción, que nos indica qué tan cerca esta nuestro ping de la velocidad de propagación teórica. Valores cercanos a 1 indican pings que se propagan tan rápido como la teoría, y valores cercanos a 0 indican pings lentos.

Latencia agregada comparada

Otra representación de las mediciones que podemos tomar es un gráfico de latencia en general, tomando todas las mediciones realizadas. De forma similar a los gráficos anteriores, dividimos las mediciones entre aquellas que son realizadas entre puntos de un mismo país y las que son entre sitios de diferentes países. Podemos ver esto, para 2017 y 2020, en el siguiente gráfico:



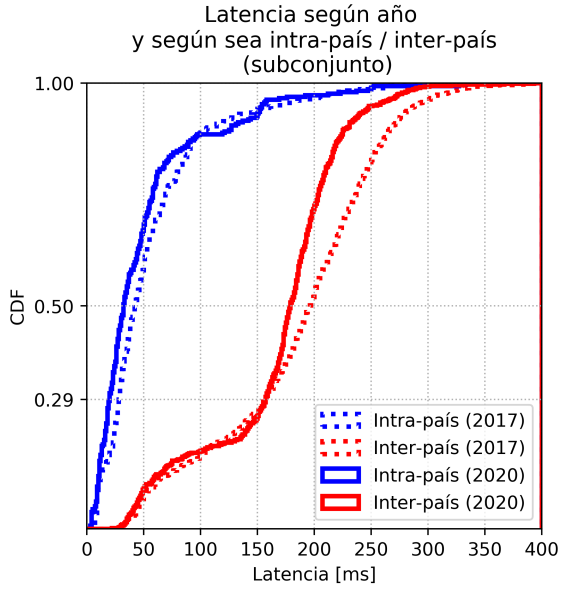
En este gráfico podemos ver las curvas de comparación entre el estudio de 2017 y el de 2020. En línea punteada se representan los tiempos de latencia en 2017 y en línea continua los de 2020. A su vez, están representadas las mediciones ***intra-país*** (es decir, entre puntos ubicados dentro de un mismo país) y las mediciones ***inter-país*** (es decir, mediciones entre puntos ubicados en diferentes países).

En el gráfico podemos ver una clara mejora en los tiempos en las conexiones internas dentro de los países. Si ponemos foco en las curvas de la izquierda, en color azul, podemos ver que la curva de 2020 (línea continua) se encuentra completamente a la izquierda de la curva de 2017, es decir representa latencias más bajas. Para 2020 el 50% de las mediciones (mediana) está por debajo de 39 ms, en comparación con 52 ms de 2017. Analizando las mediciones más lentas, el 95% de las mediciones de 2020 se encuentran por debajo de 166 ms, en comparación con 150 ms en 2017.

Sin embargo, si analizamos la conectividad entre países (curvas rojas), es decir, la latencia entre puntos ubicados en distintos países, vemos que no es tan clara la situación. Hay un porcentaje de mediciones que mejoran respecto a 2017. Podemos ver que para valores más bajos de latencia la curva 2020 es más rápida, sin embargo en el percentil 25 (122 ms en el eje de las x) ocurre un cambio y la curva 2017 comienza a ser más rápida. Sobre el percentil 72 (199 ms) la situación se vuelve a invertir y la curva 2020 vuelve a ser la más rápida. Pero la pregunta es, ¿a qué se debe ese comportamiento? ¿Estaremos cayendo nuevamente dentro de un sesgo de las mediciones?

Primero analicemos aquellas componentes que se mantienen constante entre los dos sets de datos. Por ejemplo, podemos analizar cómo es la latencia entre aquellos sistemas autónomos

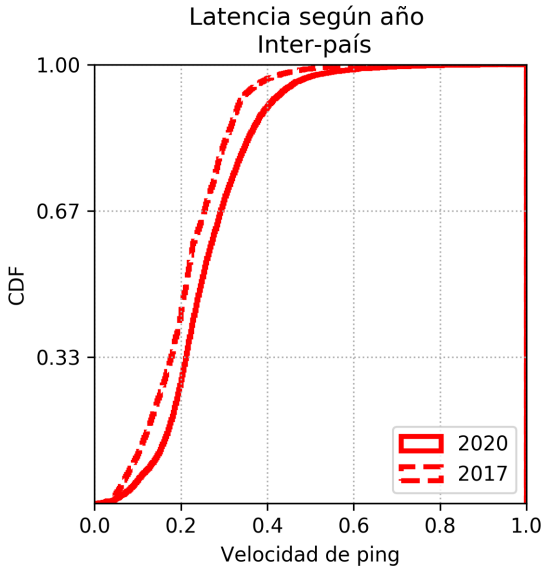
que forman parte de la campaña 2017 y a su vez de la campaña 2020. Recordemos que las mediciones tienen un sistema autónomo de origen y uno de destino, y para realizar una correcta comparación debemos mantener la tupla completa <*AS_origen, AS_destino*>. Mantener la tupla completa por un lado implica filtrar agresivamente y quedarnos únicamente con un pequeño subconjunto del set de datos original. Pero, por otro lado, trae el beneficio de que las mediciones son comparables unas con las otras, y a partir de ellas se puede inferir si ha habido un cambio real durante estos tres años. De un total de 10750 combinaciones de <*AS_origen, AS_destino*> del set de datos de 2020, nos quedamos con 373 muestras. La CDF para este subconjunto es la siguiente:



Podemos ver que para este subconjunto de mediciones, el comportamiento de la latencia intra-país se mantiene (2020 representa mejoras frente a 2017), y la latencia inter-país tiene un comportamiento un poco más claro: 2020 presenta valores menores en casi todas las partes del espectro de latencia, sobre todo a partir del percentil 29.

La conclusión del análisis de este subconjunto de mediciones es que para aquellos sistemas autónomos donde tuvimos mediciones y puntos de prueba durante estos dos años la latencia ha mejorado.

Para poder obtener mayor claridad en el comportamiento de la CDF original, repetimos el método anterior donde consideramos la distancia geodésica entre origen y destino, y comparamos la velocidad de ping frente a su valor teórico. Recordemos que esta comparación la realizamos mediante un valor que puede ir de 0 a 1, donde valores bajos indican ineficiencias en la propagación del ping, pero a medida que el enrutamiento mejora, las condiciones se acercan a lo que indica la teoría, y el valor se aproxima a 1. En este caso consideramos todo el set de datos de 2020 y de 2017. Graficando la CDF de la velocidad de ping obtenemos la siguiente figura:

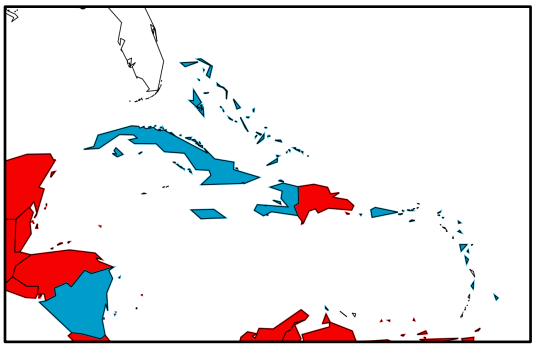


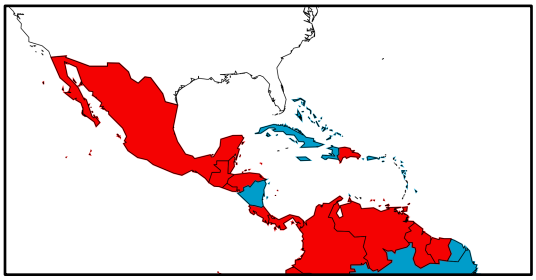
Vemos del gráfico anterior que de hecho la velocidad de ping para 2020 es consistentemente mejor que para 2017 (se encuentra a la derecha), lo que nos demuestra que han habido mejoras en la conectividad regional durante estos últimos 3 años. La curva de 2017 (punteada) se desplaza aproximadamente un 20% hacia la derecha hasta su valor de 2020. Además, gracias a esta nueva métrica, podemos concluir que el comportamiento mixto de las latencias para las mediciones inter-país es debido a la diferencia en las muestras recabadas entre un año y el otro. Una vez las muestras son normalizadas según su distancia geográfica, el comportamiento mixto desaparece.

Clusters de latencia

En base a los tiempos de latencia registrados entre países, y de la misma forma de los estudios llevados a cabo en años anteriores, agrupamos aquellos países que se encuentran más cerca uno del otro que del resto de la región.

Cuando corrimos el análisis de clusters sin supervisión, este retornó un total de 4 clusters, de los cuales 2 tienen menos de 20 mediciones, por lo que tuvimos que tener algunas consideraciones. Una vez filtrados los datos, tuvimos 2 clusters que se muestran en los mapas de abajo, cada uno en un color diferente:





Podemos ver que hay un grupo de países en la zona sur de América, que comprende Brasil, Uruguay, Paraguay, Argentina, Chile, entre otros países. Este es el cluster #1 y es el que tiene mejor interconexión, con una latencia interna de 68 ms (mediana). Estos grupos están interconectados a través de conexiones de fibra óptica desplegadas por distintos carriers que permite una alta integración. Por esa razón, los tiempos de latencia son bajos entre sí.

El cluster #0 está compuesto por la mayor parte de países de Centroamérica, algunos países de Sudamérica, y algunos países del Caribe. Este tiene una latencia interna de 129 ms (2.8 veces la del cluster #1). La definición de los dos clusters se puede ver en la imagen a la derecha.

Definición de clusters

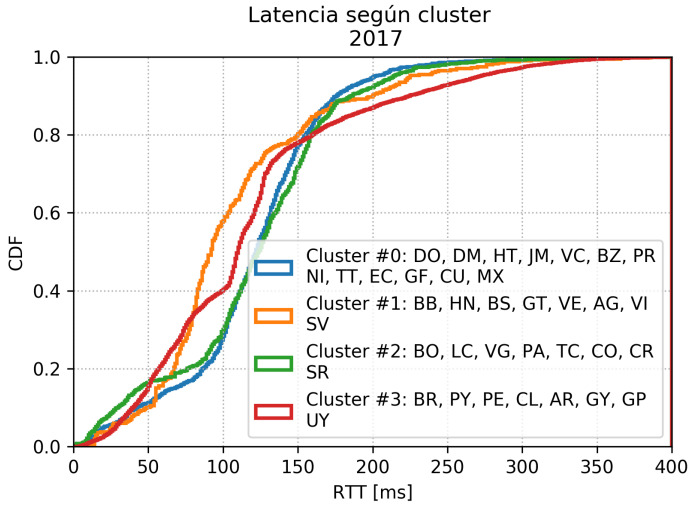
Cluster 0	**Cluster 1**
Belize	Anguilla
Bolivia	Antigua y Barbuda
Colombia	Argentina
Costa Rica	Aruba
Ecuador	Bahamas
El Salvador	Barbados
Guatemala	Brazil
Guyana	Chile
Honduras	Cuba
México	Dominica
Panamá	Guyana Francesa
Perú	Guadeloupe
República Dominicana	Haití
Suriname	Islas Turcas y Caicos
Trinidad y Tobago	Jamaica
Venezuela	Nicaragua
	Paraguay
	Puerto Rico
	Santa Lucía
	Uruguay
	Islas Vírgenes

Comparando con 2017

El mismo estudio se llevó a cabo durante 2017, del cual se obtuvieron 4 clusters.

En las siguientes dos imágenes podemos comparar las curvas de latencia en una CDF, similar a lo que vimos anteriormente. En este caso se comparan los clusters de 2017 y 2020, y las latencias corresponden a mediciones entre los países de un mismo cluster.

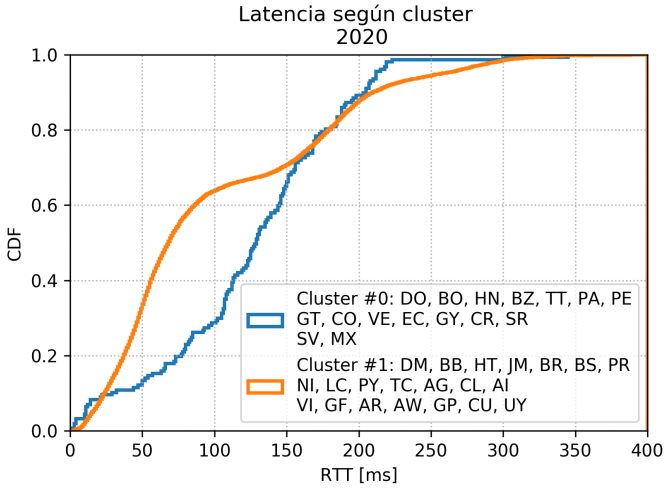
Una primera observación que podemos hacer es que en 2017 tenemos 4 clusters mientras que en 2020 tenemos 2. Analizando el caso de 2017, veíamos que hay un grupo de países en el cluster 3 que tiene baja latencia, menos de 74 ms en el 30% de los casos. Luego las latencias se incrementan, pero el 63% de las mediciones de este grupo están por debajo de 125 ms. En este grupo se encuentran países como Chile, Argentina, Uruguay, Brasil y Paraguay (ver en la figura cluster 3).



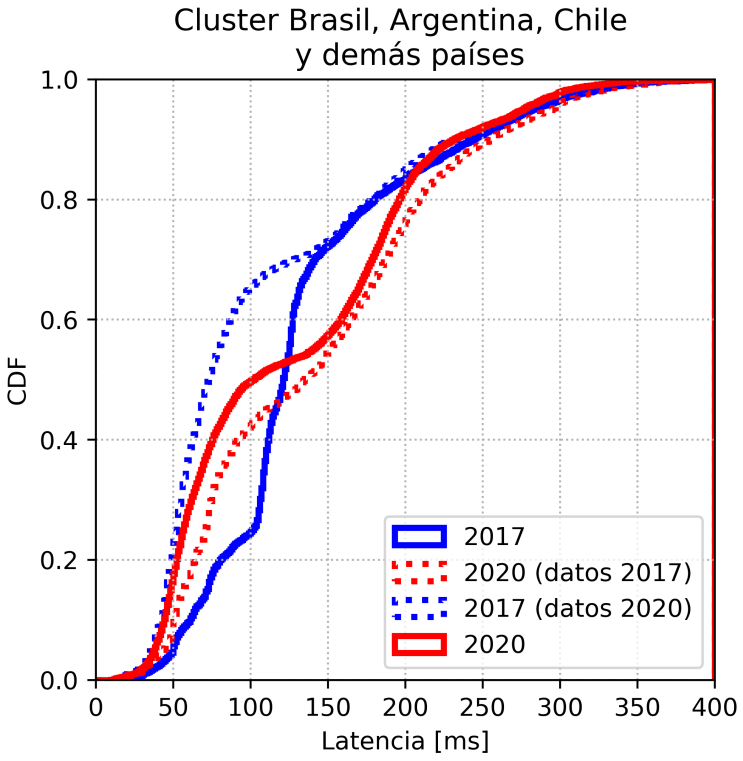
Hay un segundo grupo de países que también tiene un 73% de las mediciones por debajo de los 125 ms, e incluso en muchos casos se comporta mejor que el cluster 3. En este grupo se encuentran países como Honduras, Guatemala, Venezuela y El Salvador (ver cluster 1). En este caso, al igual que el anterior, hay un

27% de las mediciones que superan los 125 ms y que se alejan de ese valor, llegando incluso a los 300 ms.

Un tercer grupo de países compuesto por México, República Dominicana, Cuba, Belice y Trinidad y Tobago, tiene un comportamiento más uniforme: si bien por debajo de los 100 ms las latencias son levemente superiores a los clusters 1 y 3, por encima de ese valor la curva se comporta de manera similar al cluster 3, pero en este caso podemos observar que más de un 90% de las mediciones están por debajo de los 175 ms (ver cluster 0).



Por último, hay un cuarto grupo de países que está agrupado en el cluster 2, en el que se encuentran Colombia, Panamá, Costa Rica, Nicaragua y Bolivia. En este caso, las latencias están por encima de lo visto en los otros clusters, superando los 100 ms en el 70% de los casos.



Con respecto a 2020, lo primero que salta a la vista es una mejora de los tiempos de latencia en el grupo de países del cluster 1. Aquellos que estaban en el cluster 3 en el gráfico de 2017, ahora los podemos ver en el cluster 1: Chile, Argentina, Brasil, Paraguay. En este cluster aparecen más países que antes, fundamentalmente del Caribe, como Jamaica, Bahamas, Islas Vírgenes, entre otros.

Es importante notar dos cosas: la definición de los países que componen el cluster cambia de 2017 a 2020 (se suman países del Caribe), y los datos también. Por ende, podemos analizar la evolución en la latencia según dos criterios:

1. Analizando el cluster 3 de 2017, comparando los datos de 2017 con los de 2020
2. Analizando el cluster 1 de 2020, comparando los datos de 2017 con los de 2020

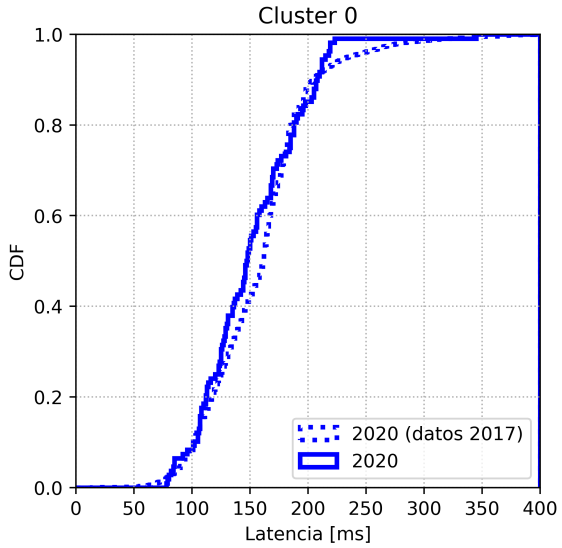
Para hacer eso agrupamos los datos bajo diferentes curvas en la CDF anterior.

Podemos ver que en ambos casos hay mejoras en la conectividad del cluster:

1. Para la definición de 2017 (línea azul vs azul punteada) vemos una clara mejora en las latencias debajo del percentil 75: el percentil 40 se desplaza de 112 ms a 63 ms, representando una mejora de 49 ms.
2. Para la definición de 2020 (línea roja vs roja punteada) vemos que las curvas tienen un comportamiento muy similar, y que se desplaza levemente hacia la izquierda. Por ejemplo el percentil 50 se desplaza de 138 ms a 103 ms, representando una mejora de 35 ms.

Esto muestra una gran mejora en la interconexión entre ese grupo de países, probablemente debido a mayores relaciones de peering y tránsito entre operadores. No obstante, al igual que en 2017, el 20% de latencias más lentas está por arriba de los 200 ms y se aleja de ese valor y llega a los 300 ms o más. Esto puede deberse a mediciones desde sistemas autónomos que se encuentran poco conectados a nivel regional y que terminan intercambiando tráfico en EEUU o en Europa.

De la misma forma, repetimos la comparación para el cluster 0 (definición 2020, datos 2020 versus datos 2017). De esta comparación no se puede decir que haya habido mejoras en los últimos años. Hay que tener en consideración que el cluster 0 no es un cluster en sí mismo, ya que agrupa todos aquellos países que no han sido agrupados en un cluster definido. Es altamente probable que a medida que se obtengan más mediciones aparezcan otros clusters con mejor latencia interna.



Conclusiones

En resumen, podemos ver que hay un grupo importante de países (que incluye algunos de los países con mayor población y desarrollo de Internet de la región) que en 2020 presentan una mejora importante respecto de 2017, donde la mediana de la latencia medida mejora en 49 ms.

Si se considera el subconjunto de redes (Sistemas Autónomos) que se midieron en ambas campañas, se observan mejoras. Además si se consideran los datos en términos relativos a las distancias geográficas (velocidad de ping), también se observan mejoras.

Otro aspecto importante es que esta campaña de mediciones de latencia incluyó dejar mediciones activas corriendo de forma permanente. De esta forma se podrá comparar, con la misma metodología, mediciones futuras de la conectividad regional. Además, a partir de esta campaña y utilizando la misma metodología, también se comenzaron a recolectar traceroutes.

Queda planteado para un futuro estudio complementar este análisis con datos sobre interconexión de redes entre este grupo de países a nivel de tablas de BGP.

Notas finales

Pueden encontrar el análisis que soporta este estudio en el GitHub de LACNIC Labs.

Las mediciones que soportan el estudio forman parte del proyecto SIMON de LACNIC. Se pueden descargar aquí.