Measurement of congestion on ISP interconnection links

David Clark
MIT CSAIL
Oct 16, 2020

Interconnection links between ISPs and other entities (such as content providers) have been a point of concern about performance degradation due to congestion. Since these links involve a bilateral agreement between the interconnected parties, failure to agree to upgrades as traffic loads increase can make these links particular points of congestion. Since these agreements involve business negotiation, and not just technical considerations, disagreements about terms can lead to overloaded links over time. As well, one of the interconnecting parties may reach a justifiable conclusion that a minor degree of occasional congestion is an acceptable outcome given a cost-benefit analysis.

Given the history of disputes over interconnection agreements, including (perhaps the most well-known) the dispute between Netflix and a number of ISPs over the terms of their direct connections, the Center for Applied Internet Data Analysis (CAIDA) at UCSD and MIT have been measuring congestion on interconnection links for a number of ISPs (including the major US ISPs) since 2016. The measurement method involves sending pairs of packets with carefully crafted TTLs to each interconnection link we have identified, one of which triggers a TTL expired from the router on the near side of the link and the other from the router on the far side of the link. If the link is experiencing congestion, a queue of packets forms (in one or both directions), and the probe packets experience increased delay as they sit in that queue. By looking for periods of increased latency from the far side of the link, we can (with some limitations) infer congestion. [1]

Since we have been collecting this data for several years now, it provides an opportunity to see if there have been significant changes since the COVID pandemic started.

Some details on the measurement and analysis

We probe each link we have identified up to three times every 5 minutes. To illustrate our method, I plot a link with recurring congestion in Figure 1 (a link between Cox and Level3 during the period of unresolved negotiations between Netflix and ISP).

In this figure, the height of the episodes indicates the size (measured in ms. of latency) of the buffer. The width indicated the duration of the congestion on each day, and is the more relevant dimension.

[1] For a full description of the method for data collection and analysis, see A. Dhamdhere, D. Clark, A. Gamero-Garrido, M. Luckie, R. Mok, G. Akiwate, K. Gogia, V. Bajpai, A. Snoeren, and k. claffy, "Inferring Persistent Interdomain Congestion", in ACM SIGCOMM, Aug 2018.

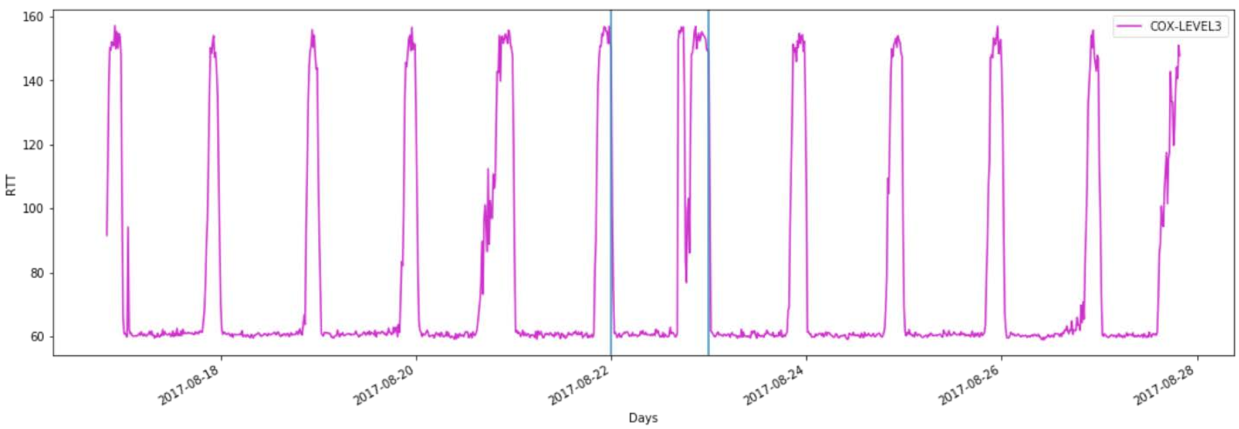

Figure 1: Recurring diurnal congestion on a link between Cox and Level3.

There can be several reasons for an episode of increased latency to the far side of a link. In our data analysis, we focus on recurring episodes (of the sort illustrated here), since the diurnal pattern give a strong hint that this is diurnal link overload, as opposed to some one-time event that is hard to classify. So we may miss some congestion events. (Work in progress aims to identify and classify some of these other events that show up in our data.)

We gather data for all links we can reach from one or more vantage points in each of the monitored networks[2], but to limit the scope of the resulting analysis, we do not analyze the links to the customers of the ISPs, but to their peers, their transit providers, and large content and cloud providers (who may, in business terms, be either revenue-neutral or paid interconnections).

Links are sometimes part of a Link Aggregation Group or LAG. We attempt to infer when LAGs are present, and the data presented here is actually for each LAG we infer, not for each link that might be part of a LAG.

For LAGs that are lightly congested, our method may miss some events. Recurring high-levels of congestion, as illustrated in Figure 1, are easy for our algorithms to detect. But if a link is congested only a few days a month (for example) we see that the timing of the events is sometimes not well correlated, and our algorithm does not find a recurring pattern. For lightly congested links, if we plot all episodes of congestion each day, rather than those that are sufficiently correlated, there are often more total events. However, in this case, evidence suggests that the periods of congestion are short.

A high-level assessment of congestion

One way to get an overall picture of the shape of congestion is to add up, for each day, the number of LAGs that experienced congestion above a certain level—lumping together peers and providers. Figure 2 shows the number of analyzed links we monitor connecting from Comcast which had more than 30 minutes of congestion on that day.

[2] Vantage points are hosted by volunteers and thus may come and go as circumstances change for those people. During 2020, we had 6 active vantage points in Comcast, 2 in AT&T, 3 in Cox and 2 in Verizon. In this paper, we have selected these four networks to evaluate.

One must view such a high-level picture with some reservations. First, a link that is included in a day might have been congested for 45 minutes or 8 hours. This rollup does not reveal that. Second, we have no way to measure the capacity of a link. A 10 gb and a 100 gb link contribute equally to the overall measure.

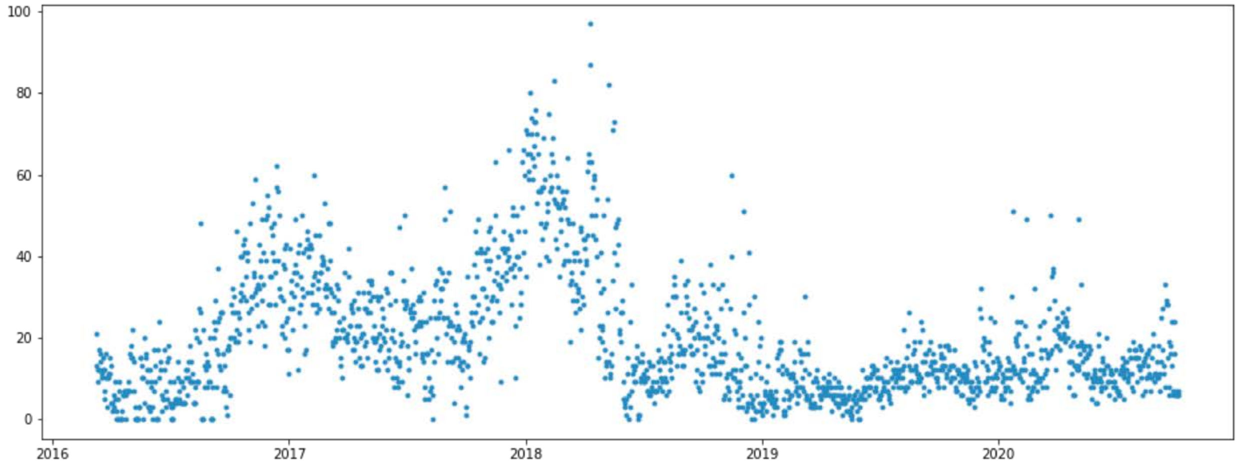

Figure 2: The daily count of LAGs from Comcast with more than 30 minutes of congestion, across all the links we analyze. The total number of analyzed links varies, but during 2020 (the period of interest with respect to COVID) was around 475.

The structure revealed by this plot shows two periods of increased overall congestion, one around the beginning of 2017 and one around the beginning of 2018. However, during 2020, there is not a consistent pattern of increasing congestion. There is a rise in the spring, which seems to coincide with the shift to working and learning at home, but this rise seems to be largely reversed. From this picture, it is hard to tell if there is a rising tendency close to the current time.

The picture reveals outlier days. I do not focus on these in this paper, because they do not seem to indicate trends, but rather result from one-time events such as the release of a new game. In our current research we are exploring these to see what can be learned, but here we focus on the larger mass of daily events.

While Figure 2 does give an overall picture, it hides a great deal of information. Most obviously, to what extent are the daily congestion events concentrated in a few interconnected parties, or are they scattered across a large number of parties. One way to look more deeply into the data is to plot the daily occurrence of congestion events for the top parties contributing to the data. Figure 3 shows this information for 2020.

What this figure suggests is that in the spring of 2020, a number of connected parties experienced an increase in congestion on their LAGs, which was corrected. The observed congestion during August and September is related primarily to LAGs associated with Akamai.

To put this data into context, each plotted number represent the total number of congested LAG-days in two weeks bins for each connected AS. Thus, a data point of 28 for a connected AS would mean that on average, that connected AS had two LAGS per day with congestion that lasted over 30 minutes in that two-week interval. We do not normalize this by the number of LAGs for the different parties, which will range from one or two to 40 or more.

The data from Akamai requires some additional background. We detected and measured 45 connection points to Akamai from Comcast, so if we detect 100 congestion events over a 14-day period, this would imply that on average, about 7 of the 45 had a congestion event on each of those days. However, our measurement method targets inter-AS links. Akamai deployed its caches much earlier than other CDNs, and for this reason, many of their caches are connected to Comcast using addresses that belong to Comcast. Since those links are not inter-AS links we do not discover and measure them. Therefore, the data we capture for Akamai in particular may not be representative of their overall levels of congestion as they connect to Comcast. I think the correct conclusion about this graph is not a specific conclusion about Akamai, but the more general conclusion that when the aggregate levels of LAG congestion are low, the shape of the aggregate data can be influenced by the behavior of a single interconnected party.

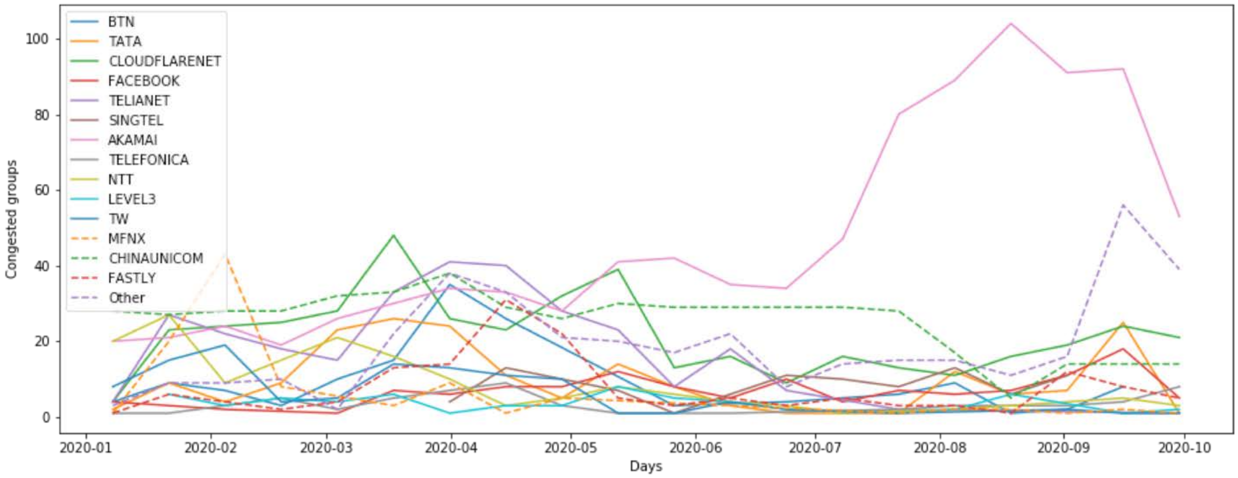

Figure 3: Contribution to overall LAG congestion from different connected parties to Comcast, aggregated in two-week bins. The remaining contributions are collected in the category “Other”. See the text for elaboration about the data about Akamai.

While the aggregate picture does mask much structure, it is often not effective to plot the entire multi-year measurement period using the more detailed method of Figure 3, because different episodes of elevated congestion are often caused by different connected parties, so the resulting picture has many jumbled lines. Here in Figure 4, for comparison, is the picture of the data by connected party over the full period illustrated in Figure 2.

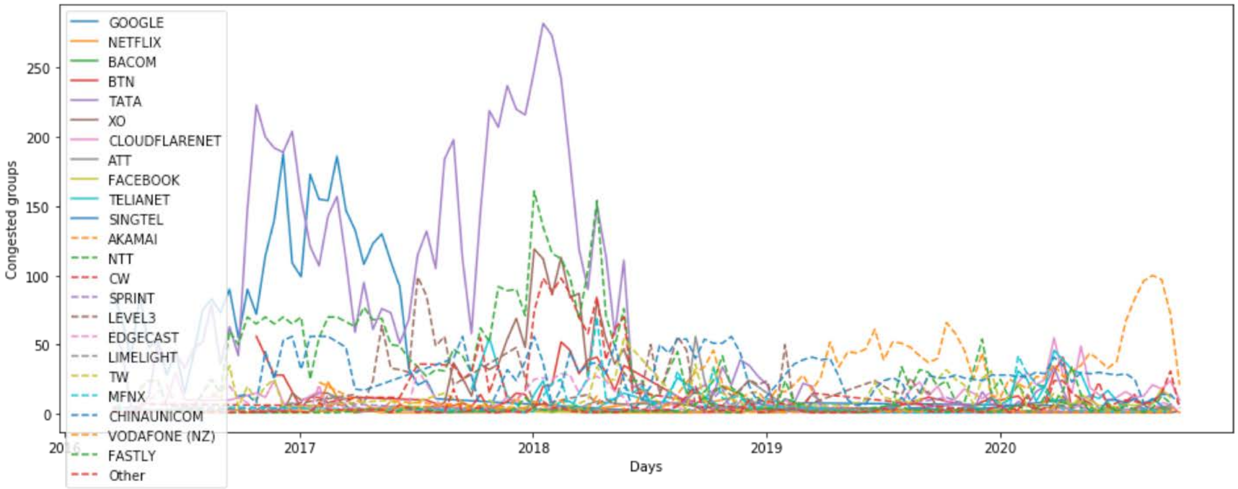

Figure 4: Contributions to congestion over the full period of data collection. The primary contributors during the early episodes were Google (Youtube) and Tata, a transit provider being used at the time as an indirect path by content providers. NTT, another transit provider, also shows elevated congestion.

Data from other US ISPs.
For comparison, here are similar plots for AT&T, Cox and Verizon.

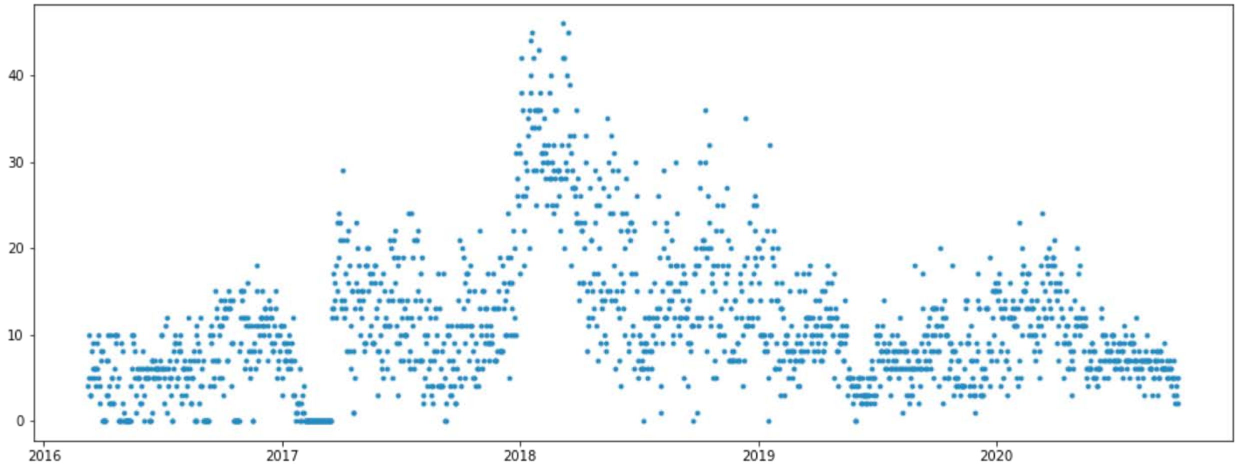

Figure 5: Aggregate data from ATT.

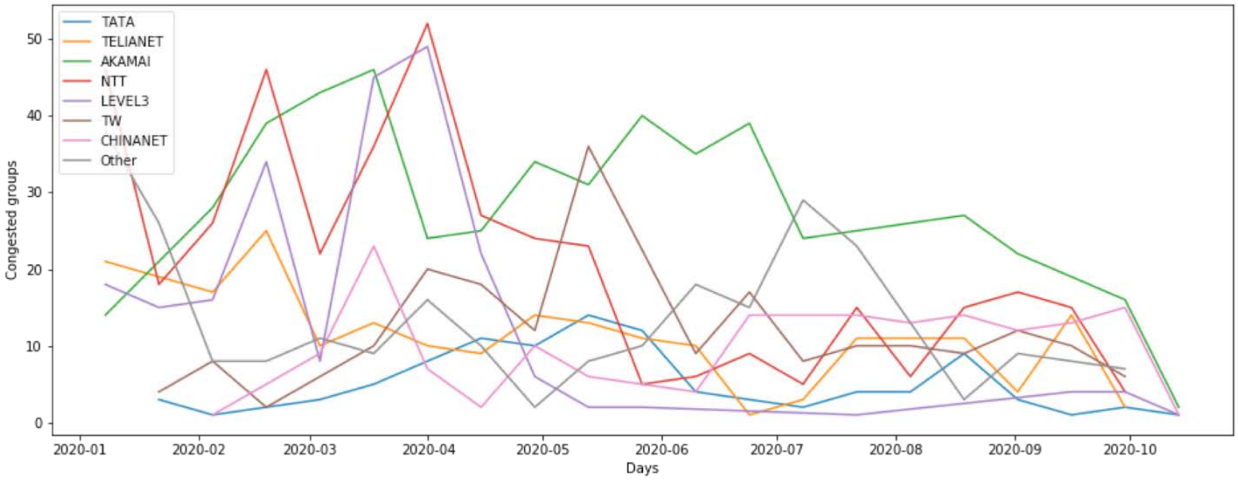

Figure 6: Contribution to overall LAG congestion from different connected parties to ATT, aggregated in two-week bins.

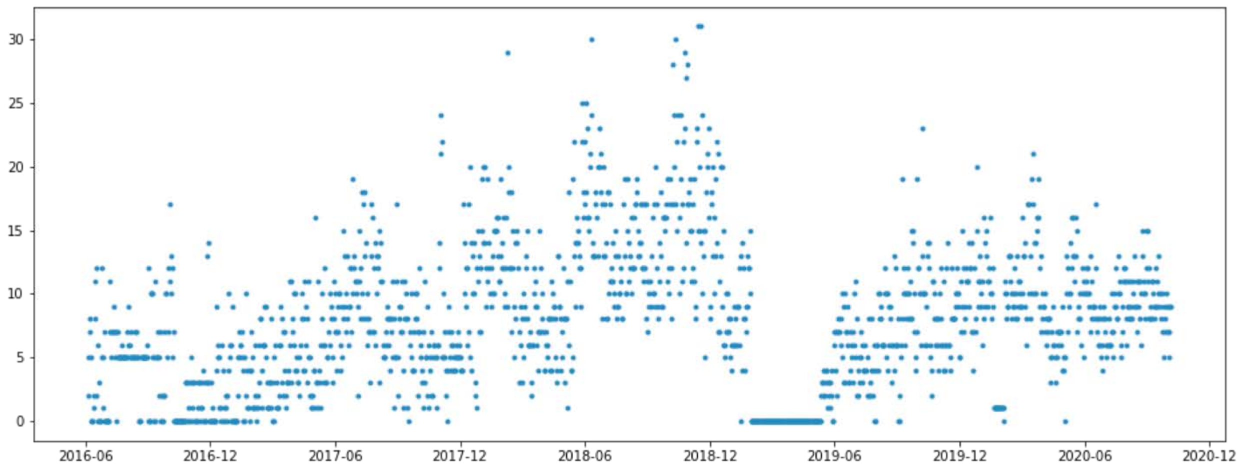

Figure 7: Aggregate data from Verizon.

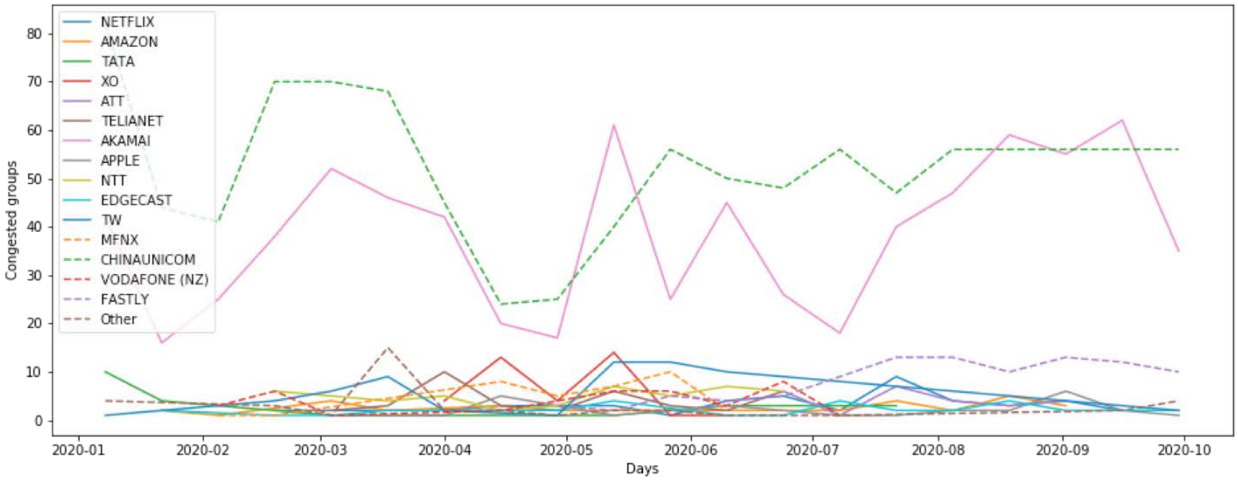

Figure 8: Contribution to overall LAG congestion from different connected parties to Verizon, aggregated in two-week bins. Note that essentially all the congestion events come only from Akamai and China Unicom.

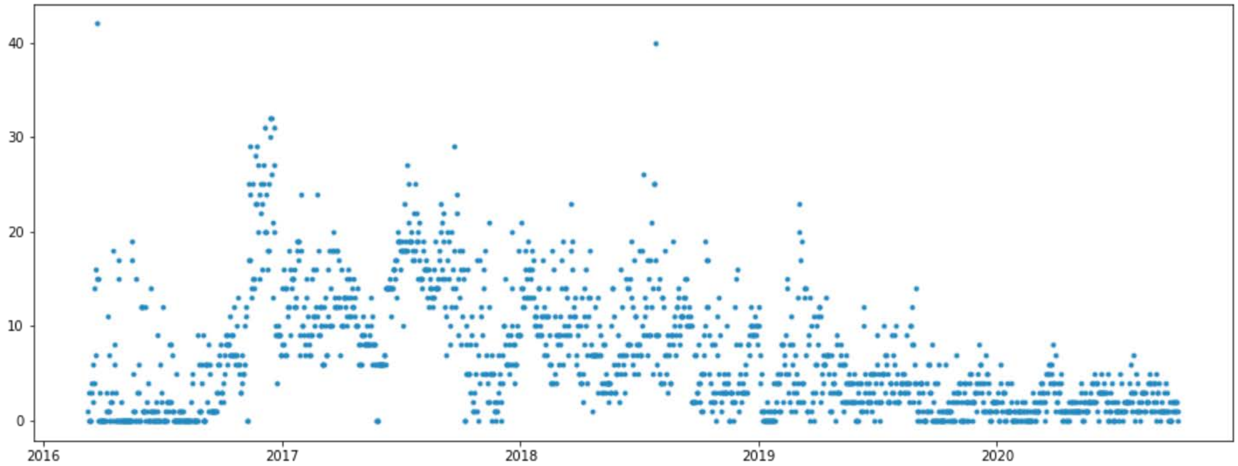

Figure 9: Aggregate data from Cox.

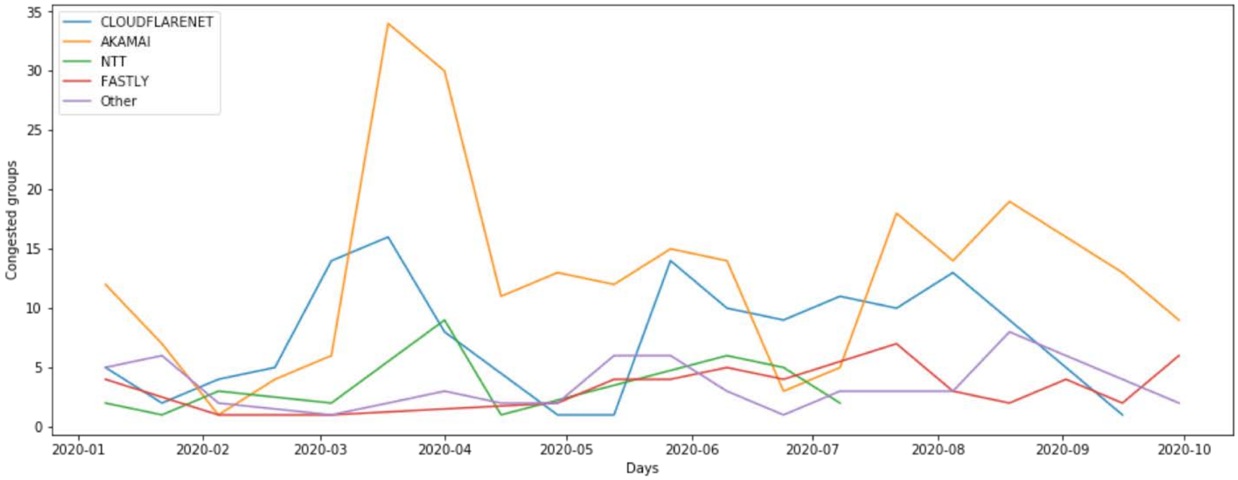

Figure 10: Contribution to overall LAG congestion from different connected parties to Cox, aggregated in two-week bins.

Conclusions

One can perhaps draw different specific conclusions from the various plots. Is there a rising trendline in some of them? And so on. I believe the correct conclusion is a higher-level point, which is both obvious and worth contemplating. Looking specifically at points of interconnection, there are no technical reasons why the Internet cannot have sufficient capacity to carry what load is offered. Adding capacity is a matter of money, planning, and agreement. (In other parts of the Internet, such as cellular systems, the technical barriers seem more substantial.) The graphs suggest that different actors have reacted in different ways to the demands imposed by the COVID crisis. This fact should not be surprising. The plots show a number of cases where congestion materialized and then vanished. This suggests that in many cases a decision was taken to allocate the necessary funding to deal with the changing demand. But what we see captured in these plots is not a technical artifact, but the consequence of economic and business artifacts.

There is a related question, outside the scope of what is measured here, which is the extent to which the slowdown and latency/jitter caused by these congestion events causes a material degradation in user quality of experience (QoE). For links that carry a known class of application traffic (such as streaming video), the provider may have an informed judgement about the degree of congestion and resulting QoE, and manage the link capacity accordingly. For other interconnection links (for example, to cloud computing), where the range of applications may be unknown and will change over time, the impact of congestion on QoE is much harder to predict.

Conversely, a provider with visibility into the link utilization of the direct links that connect it to the various ISPs may observe a link reaching a load near capacity (perhaps 80% or 85%) and start shifting to alternative sources for the content being delivered. These alternatives may not be optimal from the point of view of the application, and may cause some degradation in QoE. However, observation of the link itself, as we do, cannot detect these sorts of adaptive behavior.

Acknowledgement
The data gathered here is being collected by CAIDA, using their ARK probes and data processing infrastructure. I express thanks to the team at CAIDA that implemented and supported this work through their research and support of infrastructure, including Young Hyun, Paul Hick, Dan Andersen, Ricky Mok, Roderick Fanou, Amogh Dhamdhere (now at AWS), Matthew Luckie (Wiakato University NZ) , and kc claffy.

Funding for this work was provided by NSF OAC-1724853, NSF CNS-1513283, and DARPA Cooperative Agreement HR00112020014.