The tragedy of common bandwidth: rDDoS

Arturs Lavrenovs
NATO CCDCOE
Estonia
arturs(at)ieee.org

Éireann Leverett
Concinnity Risks
United Kingdom
eleverett{at}concinnity-risks.com

Aaron Kaplan
EC-DIGIT-CSIRC
Luxembourg
aaron[at]lo-res.org

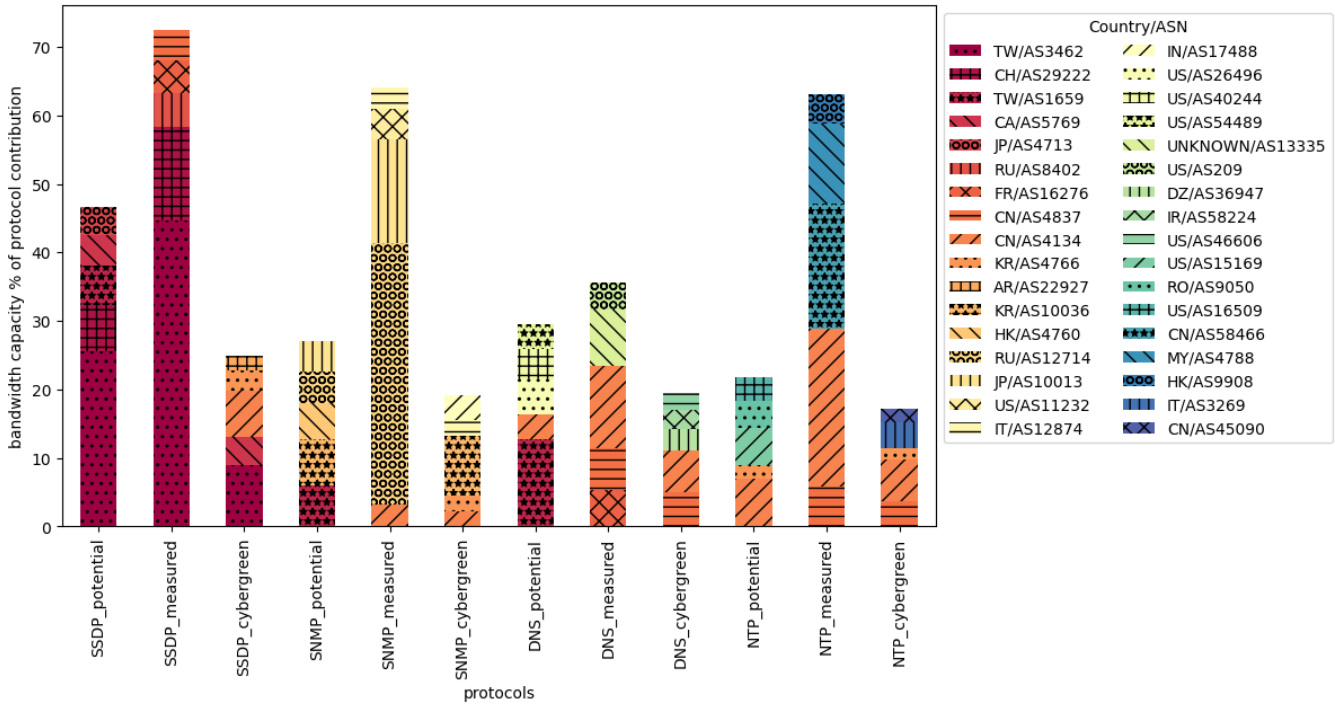

Figure 1: Biggest rDDoS "polluting" Autonomous System Names (ASNs) in May 2020 according to 3 measures (CyberGreen vs potential vs measured capacity)

ABSTRACT

Reflected distributed denial of service (rDDoS) policy interventions often focus on reflector count reductions. Current rDDoS metrics (max DDoS witnessed) favour commercial responses, but don't frame this as a problem of the commons. This results in non-objective, and non-independent discussion of policy interventions, and holds back discussion of any public health style interventions that aren't commercially motivated. In this paper, we explore multiple questions when it comes to measuring the potential for rDDoS attacks (i.e. how large could a rDDoS attack become?). We also raise some new questions. The paper builds on top of our previous research [6]. Whereas [7] was motivated by understanding properties of the individual rDDoS reflectors, in the current paper we present evidence that chasing high bandwidth reflectors is far more impactful in rDDoS harm reduction. If the internet is a commons, then high bandwidth reflectors contribute the most to a tragedy of the commons (see Figure 1). We examine and compare reflector counts, contribution estimation, and empirical contribution verification as methodologies. We also extend previous works on the topic to provide ASN level metrics, and show that the top 5 ASNs contribute between 30-70 percent of the problem depending on the protocol examined. This finding alone, motivates much easier and cheaper layered policy interventions which we discuss within the paper. The motivation of our research is also given by the surprisingly strong increase of actual (r)DDoS attacks as shown by [30]. Given this increase, our aim is to trigger policy change[1] when it comes to cleaning up reflectors. Our main contribution in this paper is to show that policy should focus on the high bandwidth reflectors and some top ASNs reduce rDDoS's potential.

NSPW '21, October 26–28, 2021, Virtual Conference

ACM ISBN 978-1-4503-8573-2/21/10…$15.00
https://doi.org/10.1145/3498891.3500928

[1] Both national as well as international public policy.

CCS CONCEPTS

• **Security and privacy → Denial-of-service attacks**; • **Networks → Denial-of-service attacks**.

KEYWORDS

Tragedy of the commons, DDoS, rDDoS, metrics, estimation, empirical, scanning, reflectors, bandwidth, interventions

ACM Reference Format:
Arturs Lavrenovs, Éireann Leverett, and Aaron Kaplan. 2021. The tragedy of common bandwidth: rDDoS. In *New Security Paradigms Workshop (NSPW '21), October 26–28, 2021, Virtual Conference.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3498891.3500928

1 INTRODUCTION

What's the biggest reflected DDoS (rDDoS) possible? Would it be of a similar magnitude if we asked the question in 2002, 2022, or 2042? Clearly, no, so what kind of magnitude scale might allow us to understand how attack severity varies over time?

Traditionally rDDoS severity has been measured by the worst that **HAS OCCURRED** historically (in Mb/s, Gb/s Tb/s, or Packets Per Second (PPS), depending on the who and when you ask). We talk regularly about the largest rDDoS on record, but ask most cyber risk professionals what the worst that **COULD OCCUR**, and they struggle to articulate a numerical answer, even if just a subjective one for a single organisation or network. Yet finding a maximum potential is necessary to disambiguate multiple causal mechanisms for the rising severity of rDDoS attacks. Without it, how might we disambiguate these potential causes for a rising trend:

(1) Bandwidth is increasing[1][2]
(2) Internet using population is growing[3]
(3) More devices are being connected to the Internet[4]
(4) DDoS defences are improving efficiency (which would imply smaller attack severity on average) [5]
(5) Some individual attacks exceed the largest previous known rDDoS event, but most do not[6]
(6) Criminals got more sophisticated in using DDoS attacks for financial extortions[7]

To clarify that disambiguation point, if you only measure attacks (and they were increasing in size) instead of also measuring potential how would you disambiguate between 1, 2, 3, and 5? To illustrate, if potential was going down attacks were still getting larger you might lean towards 5 as an interpretation. Conversely, if attacks documented were getting smaller monotonically over a few periods, how might you disambiguate 1 and 4 as causal mechanisms for the reduction in attack sizes? If 1 is getting larger, then clearly 4 is the causal mechanism for reductions (or something not yet identified).

[2]https://blog.telegeography.com/466-tbps-the-global-internet-continues-to-expand
[3]https://www.internetworldstats.com/emarketing.htm
[4]https://www.statista.com/statistics/802690/worldwide-connected-devices-by-access-technology/
[5]However, they are still the most commonly deployed answer. See Jonker, Sperotto, Rossow et. al.[30]
[6]https://www.akamai.com/us/en/resources/our-thinking/state-of-the-internet-report/global-state-of-the-internet-security-ddos-attack-reports.jsp
[7]See Jonker, Sperotto, Rossow et. al.[30]

In the history of catastrophes, measuring their impacts often begins in an ad-hoc manner, similar to simply documenting the largest rDDoS we have seen to date. Eventually though, a scale emerged, such as the Saffir-Simpson scale[2] for Hurricane wind speeds, the Richter Scale[3] for earthquakes, or the Fujita Scale[4] for tornadoes. These scales allowed further science to progress and aided experimental design that lead to better risk management. We believe it is time for rDDoS to mature and work towards a standardised scale of event measurement.

It was in this spirit of ad-hoc measurement that we began discussing what the scale of rDDoS might be. Clearly available bandwidth would be a limiting factor, since it is the very resource that becomes saturated in an rDDoS. As an overly simplistic example, you would have to be watching the biggest cables to see the maximum possible attacks. So what would those be year on year?

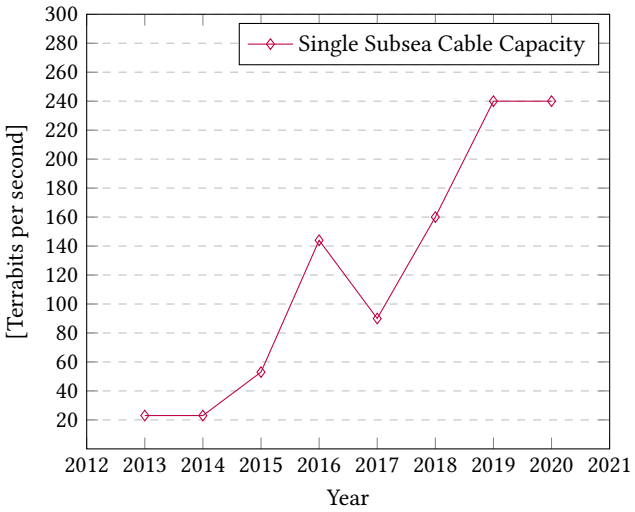

Figure 2: Maximum Capacity Cable by Year

Just because a cable could carry it, doesn't mean the available reflectors or malicious flows could produce it. To reach such a maximum number with a single event, is a max flow problem [31] of the form described on wikipedia[8]. So as a brief sanity check, Figure 2 shows the largest capacity internet cable in any of those given years[9]. Thus in 2014 and 2015, previous work on maximum estimates should be recalibrated downward to the largest cable[6], but after that, they are possible flows as a single event at least on that single cable.

This line of thinking only captures the static physical constraints, useful as that is for defining a maximum for a given point in time. However, without really discussing other mitigating factors, such as rate limiting at the device or the network perimeter, or filtering transit flows, or ASN level ingress filtering we are still missing the dynamic and defensive side of the equation. If we ignore those, we might be ignoring very significant harm reductions. Yet without a scale to begin from and thus a maximum, it is also difficult to discuss the efficacy of those interventions and mitigations.

[8]https://en.wikipedia.org/wiki/Maximum_flow_problem
[9]http://atlantic-cable.com//Cables/CableTimeLine/index2001.htm

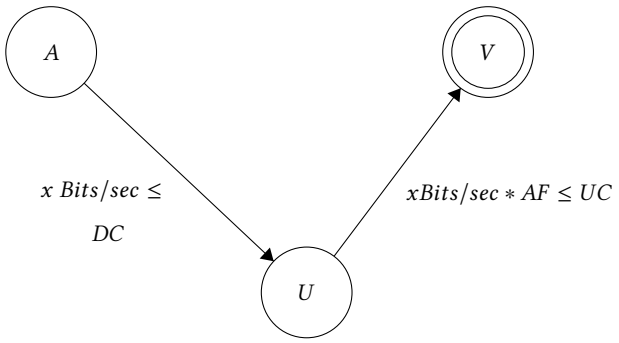

Figure 3: Model of a rDDoS reflector (amplifier)

The key issue we will continue to explore below is that reflector bandwidth and rate limiting are far more relevant than amplification factors or reflector counts. This is true for both policy and research applications, and we hope the visualisations and and text below will convince the reader that a policy and paradigm shift is necessary if we are to plan for future mitigations of maximum possible rDDoS events.

If the Internet is a commons, or a public good, then clusters of high bandwidth reflectors are polluters in a modern day tragedy of the commons[5]. The worst polluters disproportionately affecting the rest, but that also implies the promise of efficient interventions! In fact, [30] gives a clear "eye-opening statistic that one third of all /24 networks [...] have suffered at least one DDoS over the last two years"[10]. In other words, focusing on the worst polluters can have a large impact for at least one third of the internet and reduce harm for all of us.

2 RELATED WORK

Our previous paper [6] laid out the foundations for estimating the maximum possible rDDoS power. In [6], we modelled rDDoS amplification attacks, taking the upstream and downstream bandwidth of the (mis-)used reflector into account and accommodating for the mean upstream bandwidth of the CIDR netblock encompassing the reflector. See Figure 3 for a model. Furthermore, we concluded that not only the number of reflectors, but also the tier one and two (upstream) bandwidth of international carriers must be accounted for (max flow problem[31]). We also arrived at a formula for estimating the minimum rDDoS potential of the internet[11], based on the measurement of the number of reflectors and MLab's data[12] on mean upstream and downstream bandwidth per CIDR netblock.

CyberGreen data set [13] has been used in previous and current research as the source of reflector count for each examined protocol aggregated by ASN. While CyberGreen also provides its own unique *risk metric*, it is derived from the reflector count and protocol amplification factor, without considering reflecting device or network connection properties.

The model in figure 3 resulted in the following estimate for the Internet's (lower bound) rDDoS amplification power (assuming that all amplifiers are activated in parallel): the (lower bound) potential DDoS power of ASN ASN_i (in MBit/sec) is thus given by:

A	Attacker
U	UDP amplifier / reflector
V	Victim
UC	Upstream capacity (Mbits/sec)
DC	Downstream capacity (Mbits/sec)
AF	Amplification factor[14]

Table 1: Legend for figure 3

$$DDoS_{ASN_i} = \sum_{j=1}^{n} r_j(ASN_i) * \min\left(US(ASN_i); AF(r_j) * DS(ASN_i)\right)$$

r_i is the number of reflectors for the inspected protocols (SNMP, DNS, NTP, etc.) in the particular ASN. Details of this can be found in [6].

Another basis for our research is footed on empirical validation via IPv4 wide scanning[7], and exploring remediation possibilities based on this knowledge[11].

Specifically we each tried to conceive of new topic of rDDoS research: trying to measure how bad *it could be*, rather than simply recording how bad rDDoS attacks *have historically been*[30]. Unfortunately the focus of discourse around those papers has been "on the numbers"; instead of towards constructing a scale with which to measure rDDoS and global mitigations. It is not possible to overstate that bandwidth constrains the rDDoS attacks seen in practice; bandwidth constraints cap amplification factors and make the majority of reflectors a poor choice for attackers. The variance of amplification is not insignificant as discussed in the AmpMap paper[12]. Most other quantitative policy discussions revolve around reducing the reflector count or the number of incidents[13], with little discussion of severity reduction of existing attacks. We do note that lack of dialogue is at the ONLY at policy intervention level, since a healthy and efficient market of CDNs has arisen to do DDoS severity reduction. Why discuss reducing the volume of dirty bandwidth, when there are companies we can pay to clean up traffic as it flows?

We acknowledge the sophistication of the methodologies of this previous work, but question the policy framing that only attempts to reduce the number of events that actualise, and makes no attempt to estimate how many attacks could have occurred. It is the actualised rDDoS events evaluated against the background potential rDDoS events that could have occurred that motivates the authors of this paper. This would aid cyber risk estimates, in the same sense that we measure car crashes against safe journeys to get a sense of journey risk. Simplistically, we are asking which of the 5 levels of uncertainty are we at with respect to rDDoS events[14]?

Below we survey the literature to show the good work of frequency reduction and event disruption within rDDoS, before we return to these abstract concepts.

In one excellent paper we see the analysis of some natural experiments in the reduction of frequency, by 'disrupting the script' of fledgling booters and stressors[15]. Without a doubt disrupting small rDDoS attacks is good in the social benefit to deterring youngsters from crime, and reducing the numbers of events.

[10] the paper was from 2017
[11] for the protocols NTP, SSDP, SNMP and DNS.
[12] https://www.measurementlab.net/data/
[13] https://stats.cybergreen.net/download/

Other papers examine how ingress filtering can be deployed to reduce amplified flows[16][17] or how ASN accountability is a cost effective mechanism towards harm reduction[18]. This last argument is particularly persuasive given the evidence we present in this paper.

We also some see some innovative and noteworthy papers dealing with the cost estimation part of the problem[15].

Consequently, it is worth revisiting the rDDoS capacity discourse to see if new methods of estimating, validating, verifying, measuring rDDoS potential have emerged. Specifically the rDDoS potential power at the edge of the networks, before it is mitigated with middle boxes.

[30] gives the other side of the theoretical maximum. The authors created a framework based on multiple data-sets to assess how many (frequency) and how strong (intensity) actual DDoS attacks are. This view nicely complements our research.

2.1 Time as an organising principle

Comparisons of rDDoS events and capacity across time must be possible, for us to understand if interventions or risk mitigations have an impact. It is often stated that it is impossible to compare attacks in 2019 to attacks from 1999, and yet economists have learned to speak of how a 2019 dollar compares to a 1999 dollar[19][20]. Moreover, even within rDDoS, this question doesn't motivate us alone[23]:

"However, this underscores that historical attack sizes are relative, and raw numbers alone do not tell the tale. Moore's law and bandwidth increases makes comparing attack volumes (bits per second) from the past to today (or tomorrow) apples-to-oranges comparisons. Consider that gigabit attacks in 2000 were considered staggering, but only because they rivalled the capacity of the infrastructure of the time."

So, can we as an academic community, calculate an inflationary rate of DDoS (in the original depreciation sense)? If there's a Moore's law of bandwidth, might there also be one of rDDoS[21]? If we could, then it follows logically that we might be able to predict how powerful future attacks will be. This capability would allow us to both forecast the size of defences needed, as well as disambiguate what is natural growth in attack power and what is innovative attack methodology. It is a worthy goal, and not one so easily dismissed, even if early attempts to measure and define it were flawed. To critique the precision of the numbers is fair game, but we mustn't let it distract from a bigger premise of "minding the denominator" when discussing rDDoS events[22]. Thus in Section 4 we do exactly this by comparing the empirical results of one paper[7] against the estimation results of another[6].

How can we compare historical attacks against future attacks in a quantitative manner, without accounting for the growth of the internet, and innovation in attack or defence methodologies?

2.2 Effectiveness as an organising principle

When we discuss potential or estimated max rDDoS, we know that the real world is unlikely to produce attacks of this magnitude. For a start, bandwidth can impact a network flow both statically (not enough capacity) and dynamically (an integral transit link was saturated at that time). Additionally complicating the analysis; there are mitigations such as rate limiting or transit filtering. Even BCP38 might reduce the flows witnessed in actualised rDDoS attacks. When we get a recorded DDoS event, we do not get tagged data that also tells us if any of these mitigations were partially or wholly applied. It's like getting a flood report, without acknowledging that the flood defences reduced the high water mark of the actualised event.

We here advocate to see future papers include discussion of the form: *In 2021 we saw an attack that peaked at 1 Tb/s but we know it could have been 5 Tb/s if it weren't for 1Tb/s being removed from rate limiting on egress from multiple ASNs, and another 0.5 Tb/s reduction from transit filtering. BCP38 accounted for a reduction of 10 Tb/s of capacity globally, though we cannot say it affected this particular incident, as the attackers chose to use only reflectors from ASNs that do not implement it.*

In short, such metrics would motivate discussions of removing available bandwidth from attackers strategically via BCP38 or BCP84, and actualised attack severity reduction via other methods. In fact, [9] explores exactly these security economics and negative externalities in great detail. They record that only 22 percent of ASNs do not employ BCP38, which suggests that a great deal of the estimated maximum of rDDoS potential in [6] has already been mitigated at the source ASN!

Combining this information of a maximum would make ASN mitigations and their effectiveness and efficacy measurable in the large. It might even motivate a cost based implementation approach. Perhaps it would even be possible to measure a maximum rDDoS that could flow from a single ASN and get them to compete on lowering that value. Which in turn drives better cost effective mitigation strategies at the policy level.

Great examples of genre of rDDoS research are[8] or [10], and we think it would only benefit further by the addition of a scale of DDoS severity to measure harm reduction and effectiveness. In other words, they very elegantly measured the effectiveness of harm reduction against an actualised attack, but this could be extended to potential rDDoS power metrics discussed here. Only then can we examine how effective these reductions are in not just a tactical sense but also a strategic one. Or if you prefer to state this differently, how much **could be** blackholed at IXPs as measure in Tb/s reductions? What is the potential of replicating harm reduction strategies such as this much more widely across the internet?

How can we measure effectiveness, if we cannot project what would have occurred before we implemented our mitigations?

2.3 Potential reduction as an organising principle

> "Based on anecdotal data, Booter services usually can not scale up with their user base over time in terms of attack infrastructure and ultimately abandon their business at some point."[24]

This suggest some technical limit at the source/s of packet generation, or core max flow rates, as we know there certainly are enough available reflectors to go around according to CyberGreen

[15] "The impact of DDoS attacks on Dutch enterprises." a report by NBIP and SIDN

and ShadowServer[16]. Perhaps this 22 percent of ASNs analysis is showing it's value in making those reflectors useless to attackers[9].

What mitigations reduce overall rDDoS potential? Which ones give us the most Tb/s reduction for our global expenditures, either in actualised rDDoS or in potential ones[17]?

2.4 Macro and Micro interventions as organising principles

While Collier et al. [15] demonstrated the value of frequency reduction, they allow us a handy example with which to disambiguate crucially different meanings of the word *capacity*. There is the **capacity** of individuals to carry out crimes beyond their technical sophistication, such as those the NCA advertised to and disrupted within the interventions discussed in the paper. They were forced to buy DDoS from booters and stressors primarily because they couldn't organise their own. Each of those booter or stressor sites itself could be considered to have a quantitative **capacity** as well, in the sense of how many GB/s or TB/s of an attack they can launch. Presumably only one of their customers could launch an attack of that size at a time, since that capacity must be divided amongst all of their customers for single moment in time. In a directly parallel meaning we think the internet at large has a background **rDDoS capacity** (which we call potential) that must be divided between all simultaneous attacks drawing on those reflectors globally.

For us this kind of intervention research motivates the ability to estimate or empirically measure global rDDoS capacity. Thus their intervention strategies could be layered upon ours, reducing both the supply to the booters and the demand for their services at the same time.

How can we compare DDoS interventions with different methodologies and from different decades? A standardised scale will be the key to such comparisons, and since over a few decades we have not yet solved DDoS, it seems time to acknowledge a scale is needed.

3 ESTIMATED CAPACITY

In [6], we made a first attempt to estimate summed capacity in 2017 (108 Tb/s). This is not to say a single attack could reach 108 Tb/s in 2017. Rather we intended this statement to mean that all attacks that occurred simultaneously in 2017 could not have exceeded 108 Tb/s when summed together[18]. This number also ignored any middle box interventions that might reduce these attacks, and we acknowledge in practice we may never see these numbers. This is a feature of the estimate not a bug, in that they can be used as a background to measure harm reduction undertaken by rate limiting, IXP blackholing, or ingress filtering. It could be used as a denominator for attacks actualised rDDoS events.

In 2021 naturally, we need an update those numbers, and this gives us chance to articulate how the estimation methodology can be applied to any protocol found to be an rDDoS amplifier in the future. One only needs to know an amplification factor, and be able to scan for reflectors of that protocol[19].

We also extend that work here to show it can be measured empirically on a monthly basis, or for different Autonomous System Numbers, as detailed in Section 4. This allows us to reflect on the estimation methods and review them for accuracy and flaws. Scanning is the more accurate method of the two, especially at the ASN level, and thus estimation might be redundant. However, estimation is useful if you cannot scan for cost or resource reasons, such as enumerating all of IPv6.

To clarify all of this visually, see the estimated values over time in Figure 4, which also contains the max events. The difference between these two values is difficult to comprehend on a linear scale, so we zoom into it in Figure 5. We see much, much, more available rDDoS potential within the world than we think is ideal. However, this begs the question of why larger rDDoS attacks are not seen or recorded, given the availability of reflector bandwidth. The differential or ratio between these values could be a legitimate research paper in it's own right.

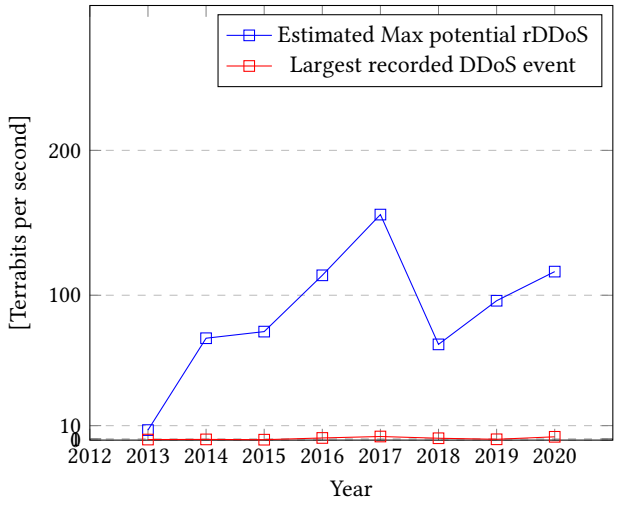

Figure 4: Available potential of reflected DDoS flows

Compare it to the peak actualised attacks in each year Figure 5, though do note the difference in scale on the y-axis. Ideally, this graph would not just show us the peak individual attacks, but the peak of summed rDDoS flows globally. This is not something we can accomplish with current data and metrics, though the large CDNs could sum concurrent rDDoS events to do so. This data is not yet possible to acquire in a cost efficient way for academics without a large collection of rDDoS honeypots, but perhaps readers can devise future methodologies to collect or analyse such data.

Especially note, how both the potential and the max documented attacks fell between 2017-2018. This is a phenomenon we cannot yet explain, but we do believe is worthy of further study. Crucially, we must ask...if the potential was reduced, could that have impacted the largest actualised events too? Or is it merely some coincidental correlation.

[16] https://www.cybergreen.net/ and https://www.shadowserver.org/wiki/

[17] We hope we're making a clear case for interventions with these top 5 ASNs, since many of them contribute more than 50 percent of malicious potential of rDDoS for a given protocol.

[18] Assuming the following protocols: NTP, DNS, SSDP, SNMP

[19] Though the scanning methodology or signatures themselves can skew the results of estimation as we discuss in the conclusion of this paper.

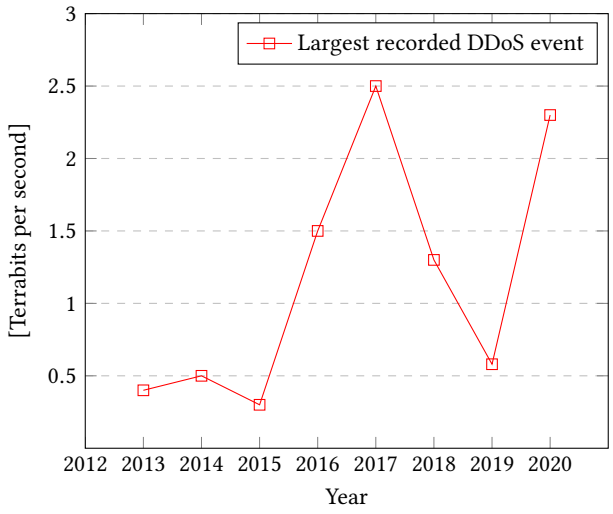

Figure 5: Actual rDDoS max events

3.1 ASN Estimates

Let's compare the top ASNs with the estimation methodology (Figure 6) against the top ASNs of empirical evidence (Figure 7). These Sankey diagrams also clarify a contribution across those ASNs and by protocol, which enriches the understanding of the top ASN contributors.Specifically we can see that the top ASNs often contribute across more than one protocol!

It is not possible to overstate the value of this insight, in that it also allows us to focus any interventions with ASNs across more than one protocol. We can also see that they are not uniform contributors, and that the top contributors often dwarf the contribution of those further down the list. These two facts alone imply a national or international policy intervention efficiency we should be keen to exploit. National telecommunications regulators take note!

Returning to a comparison of estimation and empirical methods: many of the estimated top 5 do also show up in the empirical top 5. Keep in mind you are only looking at the top 5, producing a graph of more than 5 quickly becomes unintelligible. However, similar results are found near the top of the rankings for any top N, thus justifying estimation as a cost effective method that produces similar results.

That same narrative bears out when we look at empirical data instead of estimated, though sometimes the ASN in the top N changes. This makes sense since the estimation method relied on both CyberGreen and MLab data, and mostly used average or percentiles to come to their conclusion. By empirically scanning, we avoid some of the quantitative biases inherent in the estimative approaches.

4 EMPIRICAL CAPACITY

For empirical calculations we used the measuring methodology originally proposed in [7]. It relies on extracting network metadata from responses generated by devices reachable on the Internet to which we send multiple packets requesting amplified responses. By measuring more than one packet we can avoid rDDoS honeypots and try to understand the dynamic capacity impacts discusses above.

4.1 Empirical measurement methodology

The methodology consists of two stages. First is the scanning stage in which the common Internet measurement research tool `zmap` is utilised for sending protocol specific payloads[20] to respective UDP ports[21] to all IPv4 addresses on the Internet. If the expected amplified response has been received, the second stage is executed. Second is the measurement stage where 50 (or 100 for DNS)[7] requests[22] are sent and those generating amplified responses for the specific protocol are recorded together with network metadata.

Raw measurement data can be processed to provide different views, the most important ones are bandwidth capacity, response rate limitations, bandwidth amplification factors and response timings. For this research we are combining bandwidth capacity with response rate limitations to produce bandwidth capacity in rate limited and non rate limited cases.

Speed or bandwidth contribution of a single node is measured as IP packet size divided by the time window in which all the packets from the node were received. This in itself can be variable, and thus is a form of estimation in it's own right. Therefore empirical measurement requires at least two responses to determine the time windows, without both no capacity could be calculated for an individual IP. Combined with one response from scanning at least three responses from a device are needed to get reliable capacity information, though we must allow for dropped packets. For some protocols additional capabilities are established by one or more additional requests. For this reason we exclude all the devices not meeting the minimum measuring requirements. We do this even from the set of results which are not rate limited as speed of these nodes is calculable without receiving this minimum number of packets.

Bandwidth amplification factor is an important metric for reflected amplified DDoS research. It establishes relations between spoofable Internet bandwidth (available to the attackers), and rDDoS potential capacity they can use (hosted in ASNs with poor hygiene). The estimation methodology uses an average across a set of IPv4 addresses to produce a maximum theoretical ignoring the individual differences between the nodes bandwidth. Response timings and content can indicate properties of the bandwidth or rate limiting and computing power of the nodes at those addresses. For the purposes of this research paper we simply assume that attackers have sufficient bandwidth to produce a maximum measured rDDoS, thus only the rDDoS potential bandwidth contribution is required to produce an harmful attack.

For the empirical rate limited capacity we use threshold of 80%[23] response rate. This seems arbitrary but flows logically from the fact we are working with UDP packets which do not implement

[20] Payloads generating amplified responses used in rDDoS attacks. These payloads are extracted from technical reports and reflector honeypots

[21] Ports that are associated with the abused protocols in specification or protocol documentation

[22] Request count selection is a balancing act and can be specific to every protocol. It is discussed in the previous research

[23] The produced measurement data set enables us to select any threshold, we explore these different thresholds in other research

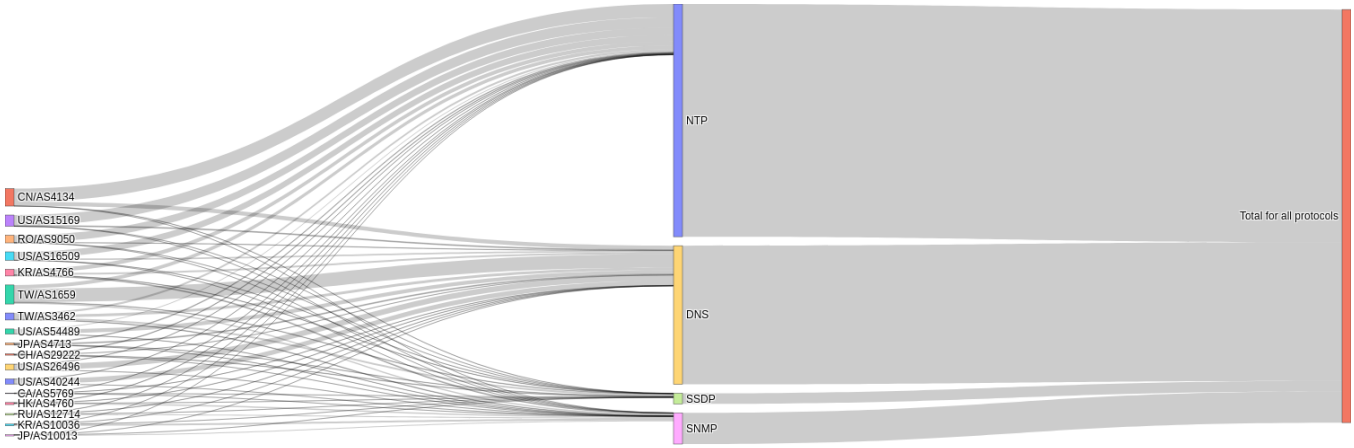

Figure 6: Top 5 ASNs by Estimated Potential in May 2020

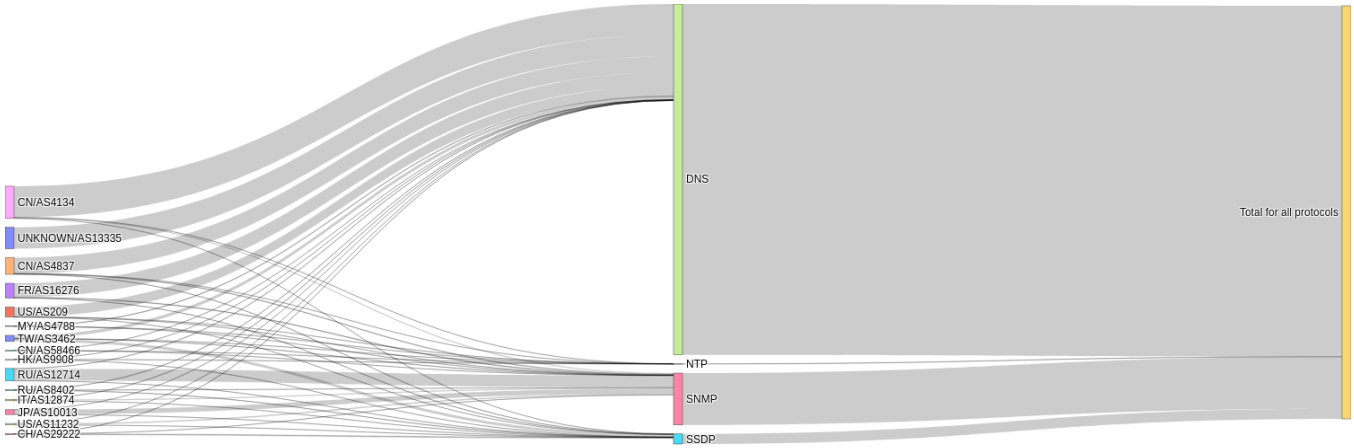

Figure 7: Top 5 ASNs by Empirical Potential in May 2020

methods to correct for lost packets. Consider our common case of 50 requests sent, we would only process nodes responding with at least 40 replies. These responses are expected within a 10 second window, or again they are excluded, which is necessary to reliably maintain receiving IP and port pair for every incoming response packet without any overlap. Thus one must choose a response threshold below 100 and above zero with which to process results. If you chose 100 per cent as a threshold many of your results would become invalid; reducing your data set considerably. If, on the other hand, you choose only 10 per cent you are reducing the number of packets you will average into an implied bandwidth contribution and lose quality there. Using 80% balances these two tensions in scientifically justifiable way. Thus we can account for the possibility of dropped packets because of network congestion or damage in transit, but also require more packets to get reliable and accurate average response rates.

See Figure 8 for a visualisation of why we use the above methodology for measurement. In particular, with NTP if you don't use such a methodology you would vastly overstate how much dangerous potential is available for NTP! Perhaps this is because NTP has been effectively mitigated over the years, and our methodology provides an excellent way to track such progress! Measuring against a estimated maximum allows us to track internet health as well as pollution. A subtle but powerful point in our quest for measurements that characterise rDDoS over decades.

4.2 Limitations

As with any Internet Measurement research there are important considerations regarding data quality[25]. For this research they are even more pronounced as individual nodes are being measured by sending fifty or more requests as opposed to common approaches of sending one or a few.

To help understand measurement limitations Figure 9 presents a simplified Internet topology with components relevant to the presented research. Victim D itself can measure an incoming DDoS attack only if it has unused incoming network link capacity and the attack isn't blocked anywhere else. Generally, these attacks have

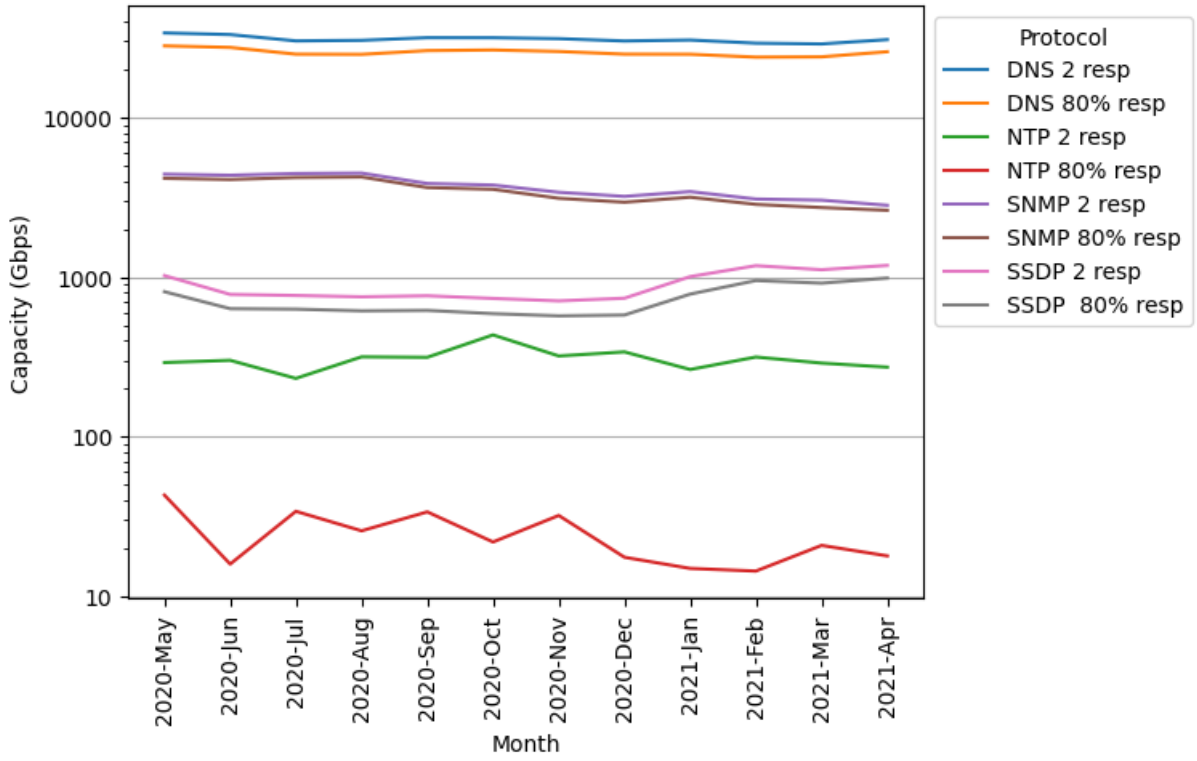

Figure 8: Empirical capacity for received 2 responses vs 80% responses

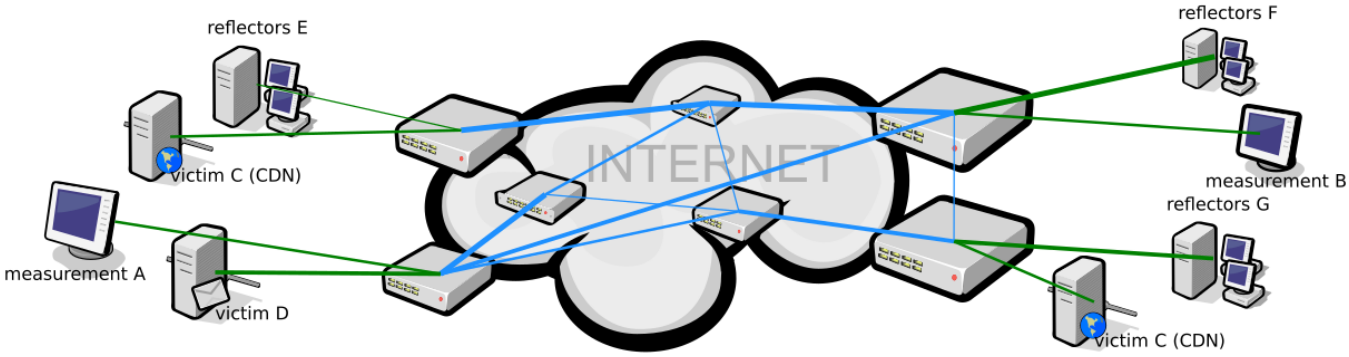

Figure 9: Simplified Internet topology with components relevant for measurements

an insignificant capacity (e.g. less than 1 Gbps on a 1 Gbps link). Larger attacks can be mitigated and measured by this victim's ISP or transit provider, the largest possible attacks often exceed mitigation ability for the majority of ISPs. Largest and record-breaking attacks can be mitigated and measured only by distributed networks that can process all incoming traffic, e.g. victim C hosted by CDN. Even this provides only the capacity of an individual attack instead of global capacity for the abused protocol.

Our measurement point A is designed to be a receiver of a DDoS attack similar to victim D, capable of receiving and processing all the incoming DDoS network traffic without external mitigation. In actuality, we are measuring maximal possible individual DDoS traffic streams and summing their contributions. The capacity we measure is specific to our position on the Internet. What would be the difference from measurement point B with the same link speed? The only guaranteed identicality is connection speed (green links) up to the closest router making decisions (e.g. BGP) for both measurement points and reflectors. Even then, the saturation of these links likely differ based on time. If that is not the case (e.g. consecutive measurements from A and B) then the difference is introduced on the Internet because of path saturation, route capacity and network policy differences (blue links).

We have verified that two different measurement points produce reasonably different capacity that maintains the difference over time. The comparison between two measurement points is usable only for network path research and establishing a capacity baseline. Over time only measurements from the same vantage point produce useful comparisons. We acknowledge that there are likely network path inefficiencies for some of the measured networks caused by the selected single point of measurement. This measurement is subjective to the vantage point and cannot be improved by averaging over multiple results. Similarly to the victim C, distributed measurement is guaranteed to produce larger capacity as packets flow shorter distances with higher bandwidth availability which contradicts our measurement objective.

Nodes that are unable to respond to all requests, because of being overloaded from the measurement are great for proving advantages of the measurement methodology. However, within this methodology, it is impossible to distinguish if a node might be overloaded for some other reason. It could be a network or processing overload because of real DDoS reflection occurring or device is using resources to fulfil its role. Then generated responses are competing for the resources and might not meet the selected threshold for being included into the overall capacity calculation.

External network issues correspond to ones that all Internet scanning activities have. Network or routing issues might occur anywhere along the path to any scanned node. While network failure is easily detectable close to the scanner it is harder to detect anomalies the closer they are to the measured node. Detecting significant temporary (shorter than the length of a single scan) anomalies is achieved by creating timeline from all the scanning and measuring metadata. Multiple scans per month allow us to exclude the ones that have the slightest suspicion to have anomalies.

Here we see why estimation approaches and empirical approaches are complimentary methods. Empirical approaches help us improve the estimation methodology, and estimations can help us "debug" when we are seeing different results because of network errors, or middle box interventions.

Temporary and permanent (across multiple scans) blocking from the target networks is a known issue. We are adhering to scanning best practice presented in [26]. Still we have identified multiple public grey lists that include our scanning IP ranges and network that utilise those. The number of unique /24 subnets per protocol that have sent at least one UDP packet within a scan or measurement stage is presented in Figure 10. The decrease in three out of four protocols include both decrease in device count and increase in network blocking. Cybergeen node count in Figure 11 presents similar picture of 3 protocols having a decrease in the node count. It might be a mix between blocking and node remediation. Without independent scan from a "clean" (not included in any grey- or blacklists) IP address range not actively used in scanning or malicious activities before it is not possible to determine how many networks have blocked our measurements.

A unique challenge that is uncommon for scanning research is DDoS defences kicking in and affecting data quality for specific target networks. It is a proper functioning of defence systems as packets we send are the ones triggering amplification and the packets we receive look like real DDoS attack on a miniature scale. It is a hard problem to address properly because of slow scan rate, churn, different scenarios of defences activating, and dynamically changing network paths. We addressed this issue minimally by selecting for the measurement node data center and transit provider that do not have automatic DDoS defences for low traffic amount.

Here too estimation can help. We propose that when a network has requested not to be scanned, a simple estimate can replace that network. Thus scans needn't exclude from results the networks that choose not to be scanned. They can either scan themselves and provide a reflector count, or we can estimate it from the distribution of reflector count per netblock.

One of the future challenges in harm reduction measurement is to distinguish between blocking[24] of scans and reflector reductions. Future research could be designed to take note of potentially blocking behaviour for (slowly) re-scanning individual suspected reflectors from a different network segment not present in any existing blacklists. Potentially statistical or mathematical techniques might derive the same answer from the data already gathered, there's no need to constrain the potential methodology to engineering solutions.

5 ESTIMATED VS. EMPIRICAL CAPACITY

Let us return briefly to Figure 1, where we hope to really clarify the difference between rDDoS severity measurement strategies. If you focus only on reflector count reductions (as represented by CyberGreen data), or only use estimation methods (as represented [6]), you would miss many opportunities for cost efficient harm mitigation. The results in this paper suggest it is possible to achieve 30-70 per cent reduction in reflective DDoS harm by working with only 5 ASNs on mitigation strategies. Let us restate that again for emphasis, by visiting any kind of incentive for mitigation on **only 5 ASNs**, from regulatory requirements to financial incentives, you could achieve volume reduction in DDoS potential of 30-70 percent in any given protocol. As if that isn't enough of a targeted intervention opportunity, the same ASNs are often heavy contributors to other protocols' pollution as well, so you needn't target 20 ASNs to achieve great impact across all the major reflective protocols. Those policy interventions could occur at the international level (Diplomatic discussions), national level (regulatory requirements), ASN level (rate limiting, reflector exclusions, BCP38), or device manufacturer level (secure by design, default configuration) to affect change in the handful of IoT device manufacturers that harm us all[27] and perhaps IoT firmware liability could be considered as an effective mechanism beyond the ASN[28]. The really key thing is that it prioritises which ASNs should really be targeted in your sphere of influence or constituency: The ones with the most bandwidth.

To drive some of those metrics differences home, let's have a look at CyberGreen's reflector node count data over time(Figure 11. Certainly great reductions have been achieved over the last twelve months in reflector counts. The variance in NTP though seems interesting. Either this suggests complications in scanning methodologies or network instability or perhaps massive variations in NTP deployment.

If their harm mitigation thesis is correct, that reducing reflector count reduces potential, then we should see such reductions in

[24] General scan blocking policy not targeting specifically us

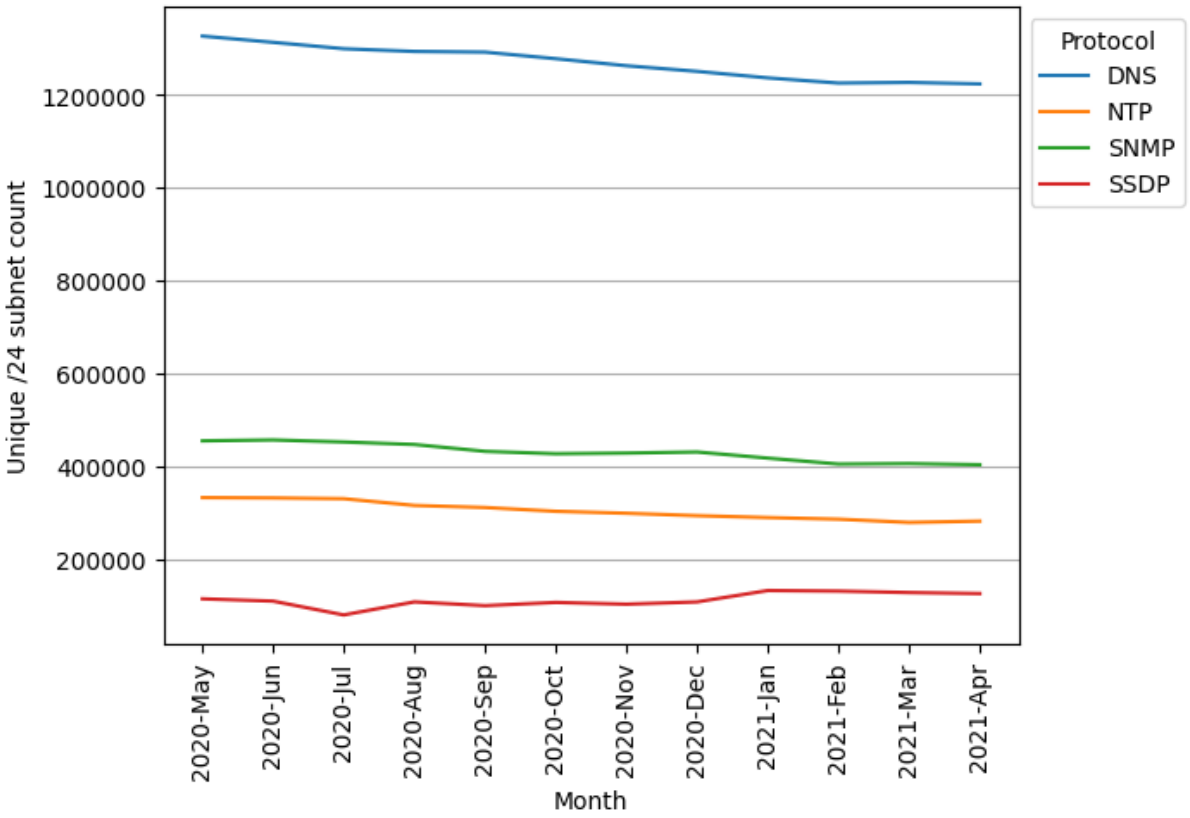

Figure 10: Unique /24 subnets per different protocols

our measured capacity (see Figure 12). Since we do not see similar reductions, we're left to conclude that the wrong reflectors are being targeted for interventions, and that bandwidth of the ASN or reflector really is the greatest contributing factor to large rDDoS attacks. To really drive this point home, notice how measured NTP potential greatly overestimates the impact (it uses CyberGreen count data), compared to the NTP measured contribution that is 3 orders of magnitude lower! Additionally SSDP reflector counts are static in Figure 11 and yet the measured potential is rising slightly in Figure 12. How is that possible if hosting bandwidth isn't the prime contributer?

Focusing on reflector count suggests interventions at many locations all around the globe are equal in priority, with relatively linear results. What we want to articulate here is that you get very non linear effects (order of magnitude improvements) by focusing on bandwidth instead of just large reflector counts. This is a fantastic opportunity in a policy sense, and a timely one given the recent documentation of carpet bombing and multi-protocol attacks[29]!

5.1 Comparing estimated and empirical capacity

Estimated capacity is faster and easier and useful if you can not scan for some reason. Empirical methods are more accurate but also more time consuming and expensive. The two are complimentary though, because the differential can tell us if mitigations are in place, or if network errors are at play, particularly when used over time. We therefore think that both have their place in organisations seeking to explore these issues.

Doing these measurements and analysis for ASNs rather than countries is the way forward. This is because the variance in bandwidth at the ASN level often lower, but also because it attributes the organisation where policy intervention might be most impactful. That in turn also removes some of the diplomatic argument that this intervention is just a tool of foreign policy, and thus are all nations considered equally responsible to focus on top high bandwidth ASNs.

We have provided country[25] based visualisations of estimated (Figure 13), empirical (Figure 14) to demonstrate these two comparative understandings (Empirical v Estimated, ASN v Country). In all visualisations (including non-geographical ones) measured values are more pronounced and extreme. Top 3 countries are the same but the first 2 places are switched, why? The first possibility could be a scanning point location in the USA for the estimated paper, and Europe for the empirical paper. This reinforces our inconvenient truth about internet measurement about how we get differing results from where and when we scan. Other discrepancies

[25] Countries with insufficient data are greyed out

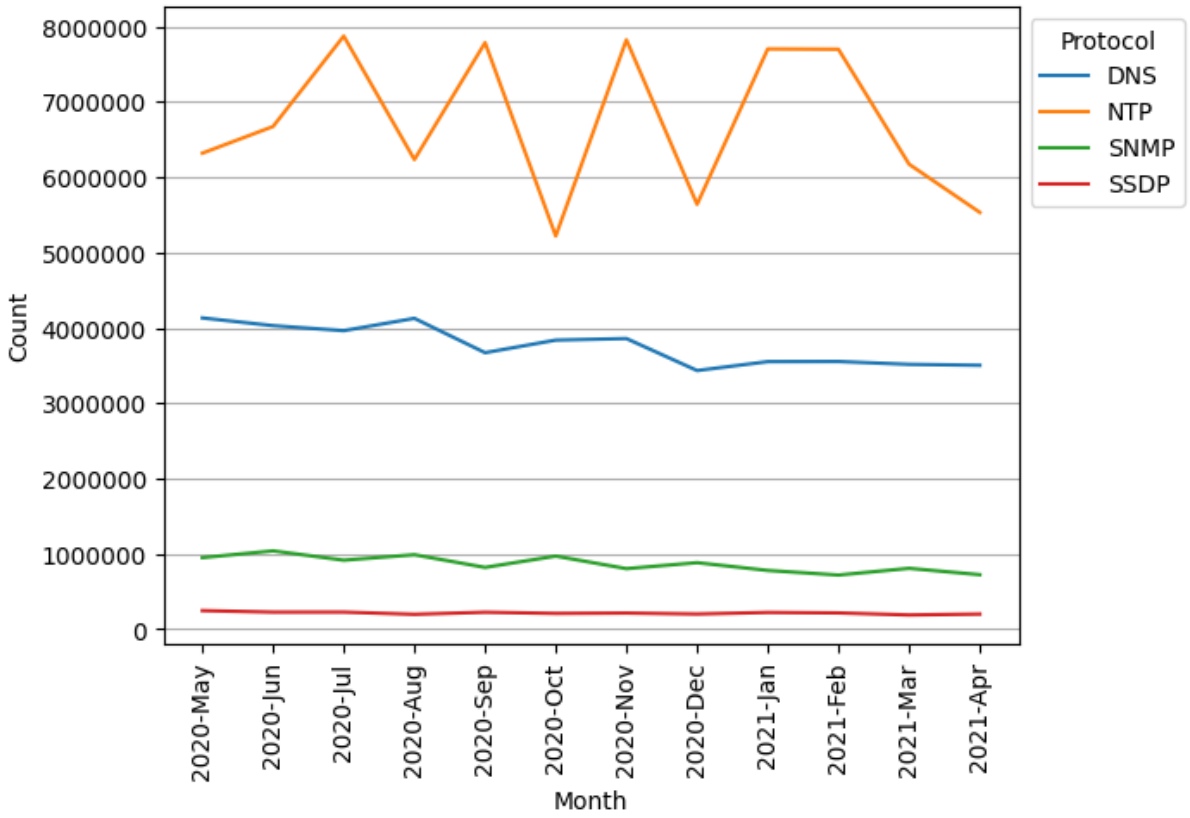

Figure 11: CyberGreen node count

between the papers could be explained by differences in bandwidth calculation and detected rate limiting.

While these are great representations of summarised bandwidth capacity across all protocols we have identified that individual protocol considerations raise additional research questions. Taiwan has the highest both empirical and estimated capacity globally for the SSDP protocol. South Korea has the highest estimated but a low empirical capacity for the SNMP protocol. It doesn't mean these countries have the worst reflector problem...quite the opposite can be true if rate limiting and source address verification are taken into account. All protocol visualisations demonstrate that these are average countries in absolute and relative terms. It might indicate that some ISPs deploy devices with a default configuration running these abusable protocols that is also unique for these countries. In the SSDP case these protocols might not have a rate limiting or have a higher one than the measurement uses. SNMP is a more interesting case as potential indicates that there is a significant number of abusable devices but measurement indicates that these devices are either rate limited or bandwidth limited and therefore less contributing to the real global attack capacity.

This is merely one example, to illustrate that both country, asn, and even protocol specific considerations all contribute in unique ways to our metrics. We must be careful to communicate the implications of those factors on the research. Plenty of opportunity for future research examining the interplay of those factors on metrics and these issues.

To illustrate this interplay a little further we have also produced four geographical representations which normalise relative to human population (Figures 15 and 16) or IP assignment (Figures 17 and18). The Internet propagation differs vastly across the globe, the allocation is disproportional. We are focusing here on IPv4 ranges which are mostly allocated to countries with early Internet adoption. The per capita representation is not ideal as a visualisation, though it is still important. Even with allocation the Internet availability varies vastly within countries. As we are concerned only with reachable devices, developing countries with limited IPv4 address allocations might employ NAT or similar solutions. In the case of internally spoofed traffic or IPv6 direct assignments this capacity might increase but that is currently indeterminable for technical reasons (to our knowledge).

We hope that seeing these visualisations makes it clear that ASN level discussion are far more accurate quantitatively, and also neutralise some of the diplomatic discussions. For example, we have heard it said that rDDoS is caused "by all those reflectors in Africa", and wish to rebut that point with "bandwidth in the global south is often much reduced". Thus regardless of large reflector counts the global south do not contribute nearly as much to DDoS pollution as developed nations.

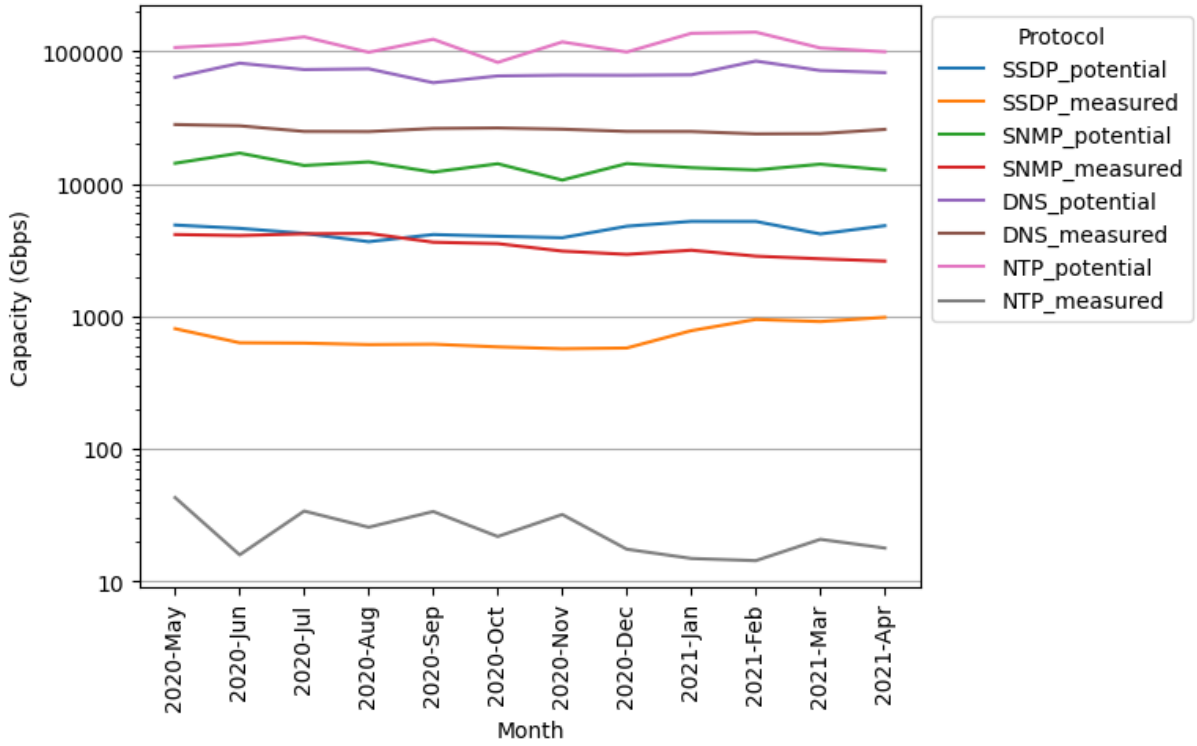

Figure 12: Potential vs measured bandwidth capacity

6 CONCLUSIONS

The internet is a spectacular example of a public good, and bandwidth in particular enables many social benefits in a cast variety of topics. rDDoS traffic in particular is a form of pollution on the commons; malicious use of a public good. As we have detailed above, multiple mitigation approaches are possible and complimentary. At the core of rDDoS though is the capacity of reflectors to harm, and by reducing this capacity at around 20 ASNs per year, we can significantly reduce the severity of the harm. Interventions with booters and stressors will continue to reduce the frequency. In combination these two strategies can "mindfully herd" the remaining malicious actors to fight over less and less resources.

The estimation methodologies in [6] produced good insights that bandwidth was more important than reflector count. Unfortunately, different scanning methodologies can skew the results of the estimation significantly. Thus any scanning methodology must establish amplification and packets received to provide more accurate estimation. In [7] a methodology for doing so is laid out, and we expanded upon it here. The two together are complimentary approaches that may lead us towards a more timeless scale of measuring DDoS.

Ideally that scale:

(1) Allows comparison of peak flow events across decades
(2) Allows duration of events to be characterised
(3) Allows us to compare what occurred to what could have occurred
(4) Allows normalisation across population or announced IP addresses
(5) Future proofs the metrics against future internet addressing schemes
(6) Allows easy translation into the cost of attack and cost of mitigation for risk calculations

With respect to empirical measurement, we have observed that individual devices have different properties. But even abstracting to protocol level it becomes clear that not all protocol implementations are built equally. Measuring different protocols has to be accomplished differently, and yet still this individuality should not skew the metric we use to measure the scale of the DDoS attack overall.

Ideally future research in this vein would tackle a few key problems. One is producing a scale of DDoS events that satisfies the above requirements. Simply stating that attacks are bigger without detrending 2008 bandwidth from 2018 bandwidth is insufficient. Secondly, it might tackle how to use backtrace methods to determine which reflectors are the most widely used in malicious attacks. This would be another way to validate or invalidate our claim that bandwidth is more important than reflector count for mitigation. Thirdly gathering data that allows us to measure concurrent rDDoS events, even where the victims are different addresses. This last one may seem obscure and strange, but it would aid us in producing a sense of what the maximum rDDoS events could be.

That in turn would aid in making a scale, and then building models of DDoS cyber risk that would allow us to evaluate policy decisions around defences.

From a policy perspective, there is much to learn from our work. The key takeaway is that it is the bandwidth rich who enable reflector and amplifier abuse, and very unevenly so. This in turn suggests that only a small number of policy interventions with a handful of ASNs could greatly clean up the commons of the internet. It is also possible to target the device manufacturers with limited liability in case of their use in DDoS events. Alternatively one might assign liability partially to the ASNs who contribute bandwidth as well. If the devices with abusable protocols are the fire, the bandwidth is the gasoline poured on the problem.

There are also policy implications for metrics research and harm reduction feedback mechanisms. One is that in internet measurement data quality is always open for discussion, and should be. We must adapt to data quality issues, and acknowledge how much they skew our estimations. We can also confidently say there is a place for empirical and estimated measurement methods, and that the two compliment each other. ASN level measurement is advantageous because it is also conducted at a scale where some responsibility can be assigned. Countries are too broad of a bucket, and the variation within them for reflector count, bandwidth, and allocated address can be to high to make insights easy.

In conclusion, if we want clean bandwidth for the future, we must make some measurements to enable it. We must hold multiple stakeholders to account, from the device manufacturer, to the device owners, to the hosting provider. The mitigation in the middle have become highly effective, but we need ways of measuring their success so we can further incentivise them. For a cleaner safer internet, we must be careful how we scan and assign responsibility for action. Let those who contribute the most to rDDoS attacks do more of the work to reduce them. Incentive engineering is hard, but doubly so if you focus on the wrong numbers. There are many millions of reflectors in the world, but a very small fraction of them contribute the most to the problem. This is the tragedy of dirty bandwidth.

REFERENCES

[1] Prasad, Ravi, Constantine Dovrolis, Margaret Murray, and K. C. Claffy. "Bandwidth estimation: metrics, measurement techniques, and tools.", in IEEE network 17, no. 6 , 2003, pp. 27-35.

[2] Saffir-Simpson hurricane wind scale, https://www.nhc.noaa.gov/aboutsshws.php, Saffir-Simpson Hurricane Wind Scale, NOAA, Accessed: 13/05/2021

[3] Richter Scale, https://simple.wikipedia.org/wiki/Richter_scale, Wikipedia, Accessed: 13/05/2021

[4] Fujita Scale, http://www.ehso.com/disasters_scale.htm, EHSO, Accessed: 13/05/2021

[5] Hardin, Garrett. "The tragedy of the commons." Journal of Natural Resources Policy Research 1, no. 3, 2009, pp. 243-253.

[6] Eireann Leverett, Aaron Kaplan. "Towards estimating the untapped potential: a global malicious DDoS mean capacity estimate", Journal of Cyber Policy, 2:2, pp. 195-208, 2017, DOI: 10.1080/23738871.2017.1362020

[7] Arturs Lavrenovs. "Towards Measuring Global DDoS Attack Capacity." 11th International Conference on Cyber Conflict (CyCon 2019), Estonia, 2019.

[8] Dietzel, Christoph, Anja Feldmann, and Thomas King. "Blackholing at ixps: On the effectiveness of ddos mitigation in the wild.", in International Conference on Passive and Active Network Measurement, pp. 319-332. Springer, Cham, 2016.

[9] Lone, Qasim, Maciej Korczyński, Carlos Gañán, and Michel van Eeten. "SAVing the Internet: Explaining the Adoption of Source Address Validation by Internet Service Providers.", in Workshop on the Economics of Information Security, 2020.

[10] Kopp, Daniel, Christoph Dietzel, and Oliver Hohlfeld. "DDoS Never Dies? An IXP Perspective on DDoS Amplification Attacks.", 2021.

[11] Arturs Lavrenovs. "Towards Remediating DDoS Attacks.", in 16th International Conference on Cyber Warfare and Security (ICCWS 2021), Tn, USA.

[12] Moon, Soo-Jin, Yucheng Yin, Rahul Anand Sharma, Yifei Yuan, Jonathan M. Spring, and Vyas Sekar. "Accurately measuring global risk of amplification attacks using ampmap.", in 30th USENIX Security Symposium (USENIX Security 21), 2021.

[13] D. R. Thomas, R. Clayton and A. R. Beresford, "1000 days of UDP amplification DDoS attacks.", in 2017 APWG Symposium on Electronic Crime Research (eCrime), 2017, pp. 79-84, DOI: 10.1109/ECRIME.2017.7945057.

[14] Spiegelhalter David J. and Riesch Hauke. "Don't know, can't know: embracing deeper uncertainties when analysing risks", in Phil. Trans. R. Soc. A.3694730–4750, 2011.

[15] Collier, Ben, Daniel R. Thomas, Richard Clayton, and Alice Hutchings. "Booting the booters: Evaluating the effects of police interventions in the market for denial-of-service attacks.", in Proceedings of the Internet Measurement Conference, pp. 50-64, 2019.

[16] Zhauniarovich, Yury, and Priyanka Dodia. "Sorting the Garbage: Filtering Out DRDoS Amplification Traffic in ISP Networks.", in 2019 IEEE Conference on Network Softwarization (NetSoft), pp. 142-150. IEEE, 2019.

[17] Du, Ping, and Akihiro Nakao. "DDoS defense deployment with network egress and ingress filtering.", in 2010 IEEE International Conference on Communications, pp. 1-6, IEEE, 2010.

[18] Simon, Daniel, Sharad Agarwal, and Dave Maltz. "AS-based accountability as a cost-effective DDoS defense.", 2007.

[19] Bryan, Michael F. "On the Origin and Evolution of the Word Inflation.", Federal Reserve Bank of Cleveland, Economic Commentary, 10.15.1997.

[20] Dornbusch, Rudiger W. "Purchasing Power Parity", March 1985. NBER Working Paper No. w1591, Available at SSRN: https://ssrn.com/abstract=336331

[21] C. A. Eldering, M. L. Sylla and J. A. Eisenach. "Is there a Moore's law for bandwidth?", in IEEE Communications Magazine, vol. 37, no. 10, pp. 117-121, Oct. 1999, DOI: 10.1109/35.795601.

[22] Jardine, Eric. "Mind the denominator: towards a more effective measurement system for cybersecurity.", in Journal of Cyber Policy 3, no. 1, 2018, pp. 116-139.

[23] Osterweil, Eric, Angelos Stavrou, and Lixia Zhang. "21 Years of Distributed Denial-of Service: Current State of Affairs.", Computer 53, no. 7, 2020, pp. 88-92.

[24] Karami, Mohammad, and Damon McCoy. "Understanding the Emerging Threat of DDoS-as-a-Service.", In LEET, 2013.

[25] Bano, Shehar, Philipp Richter, Mobin Javed, Srikanth Sundaresan, Zakir Durumeric, Steven J. Murdoch, Richard Mortier, and Vern Paxson. "Scanning the internet for liveness.", in ACM SIGCOMM Computer Communication Review 48, no. 2, 2018, pp. 2-9.

[26] Z. Durumeric, E. Wustrow, and J. A. Halderman, "ZMap: Fast Internet-wide Scanning and Its Security Applications.", presented as part of the 22nd USENIX Security Symposium (USENIX Security 13), Washington, D.C., 2013, pp. 605–620.

[27] Rodríguez, Elsa, Arman Noroozian, Michel van Eeten, and Carlos Gañán. "Superspreaders: Quantifying the Role of IoT Manufacturers in Device Infections."

[28] Leverett, Eireann, Richard Clayton, and Ross Anderson. "Standardisation and certification of the 'Internet of Things'.", in Proceedings of WEIS, vol. 2017, 2017.

[29] Heinrich, Tiago, Rafael R. Obelheiro, and Carlos Alberto Maziero. "New Kids on the DRDoS Block: Characterizing Multiprotocol and Carpet Bombing Attacks.", in PAM, pp. 269-283, 2021.

[30] Mattijs Jonker, Alistair King, Johannes Krupp, Christian Rossow, Anna Sperotto, and Alberto Dainotti. 2017. "Millions of targets under attack: a macroscopic characterization of the DoS ecosystem.", in Proceedings of the 2017 Internet Measurement Conference (IMC '17). Association for Computing Machinery, New York, NY, USA, 100–113. DOI: 10.1145/3131365.3131383

[31] Ford L.R., Fulkerson D.R. "Maximal Flow Through a Network.", in Gessel I., Rota GC. (eds) Classic Papers in Combinatorics. Modern Birkhäuser Classics. Birkhäuser Boston, 2009. DOI: 10.1007/978-0-8176-4842-8_15

Appendix A GEOGRAPHICAL IMAGES

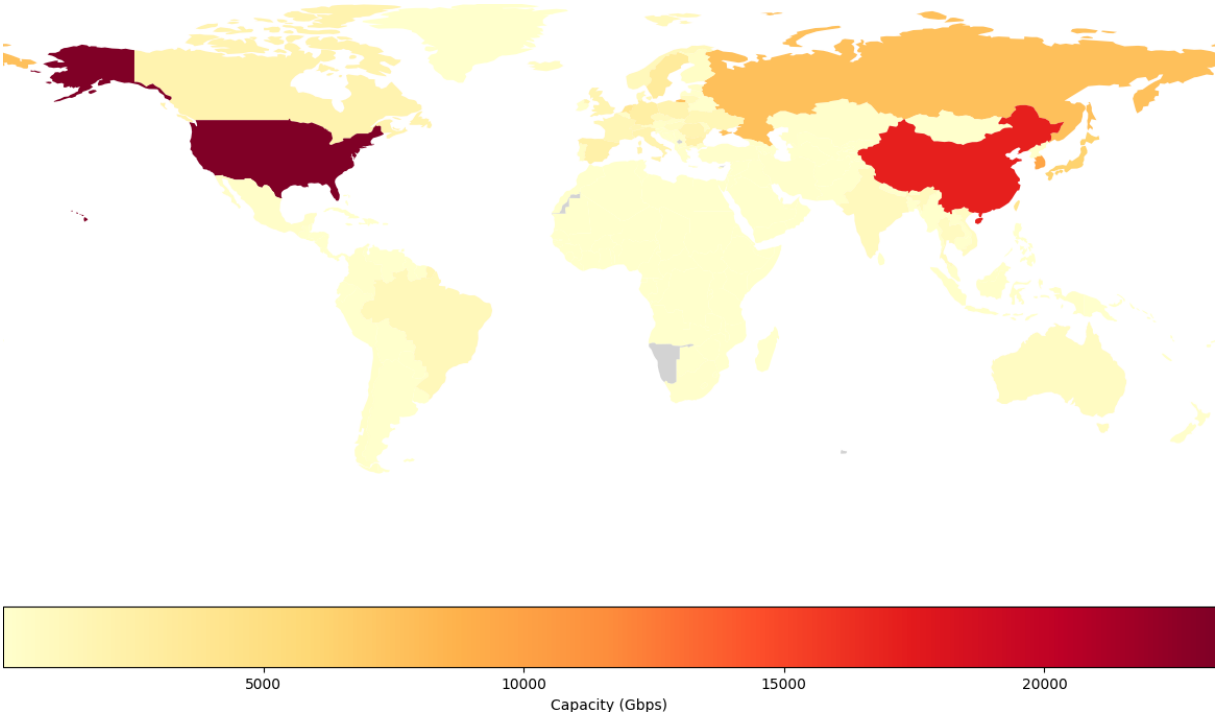

Figure 13: Potential capacity for DNS, NTP, SSDP, SNMP protocols in May 2020

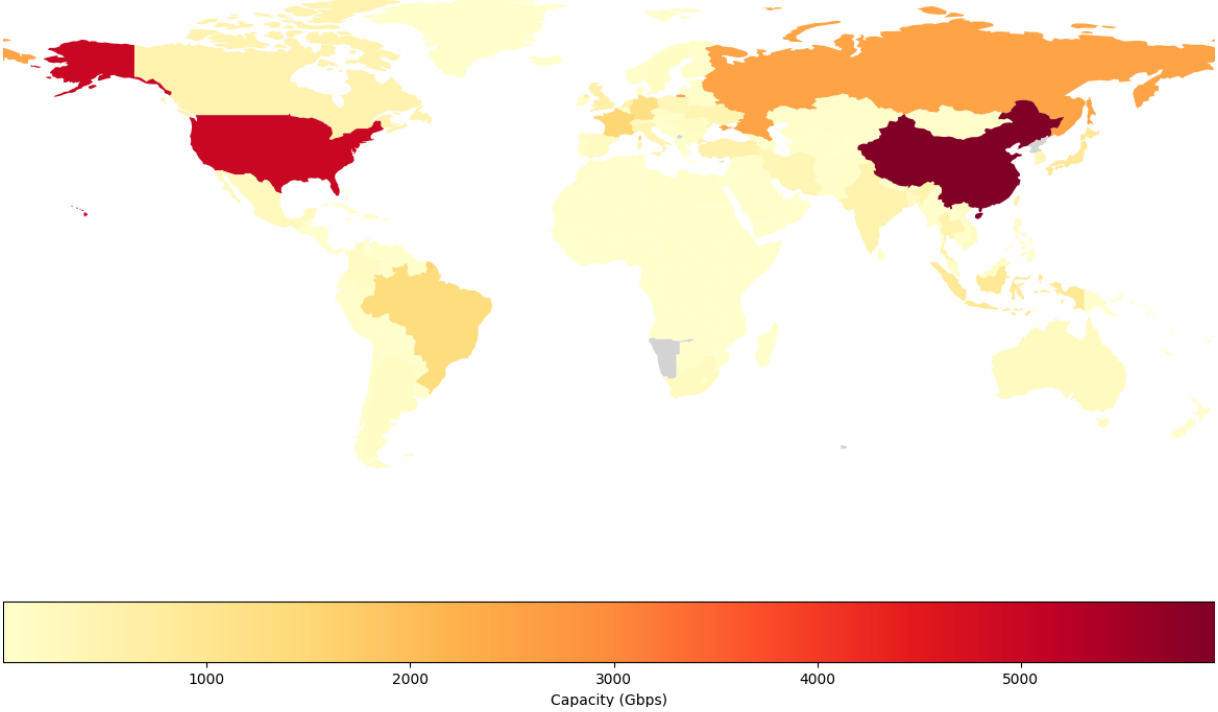

Figure 14: Measured capacity for DNS, NTP, SSDP, SNMP protocols in May 2020

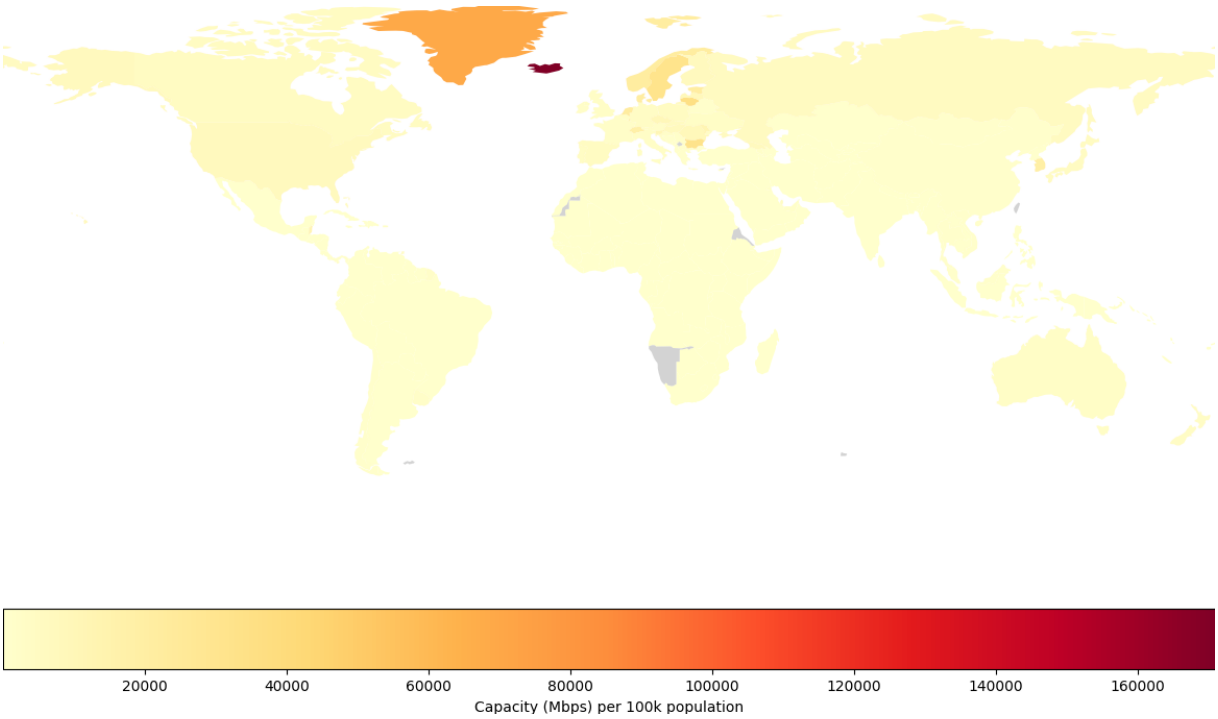

Figure 15: Potential capacity relative to population in May 2020

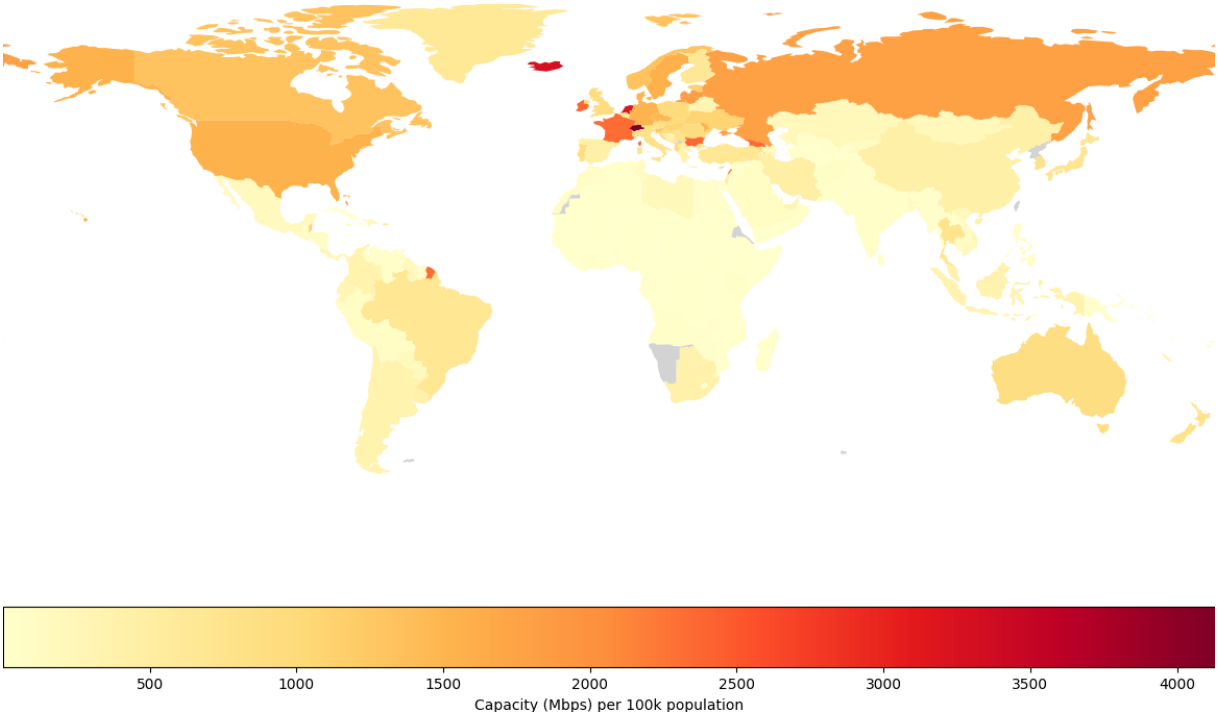

Figure 16: Measured capacity relative to population in May 2020

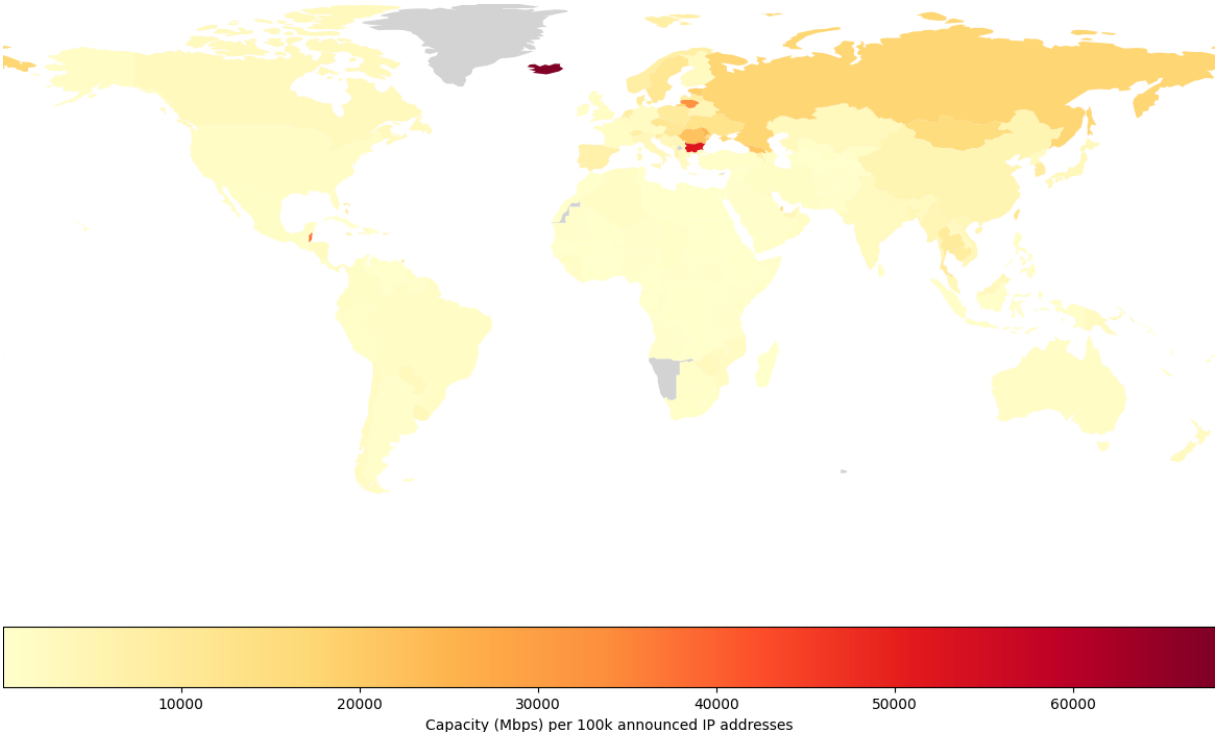

Figure 17: Potential capacity relative to announced IP addresses in May 2020

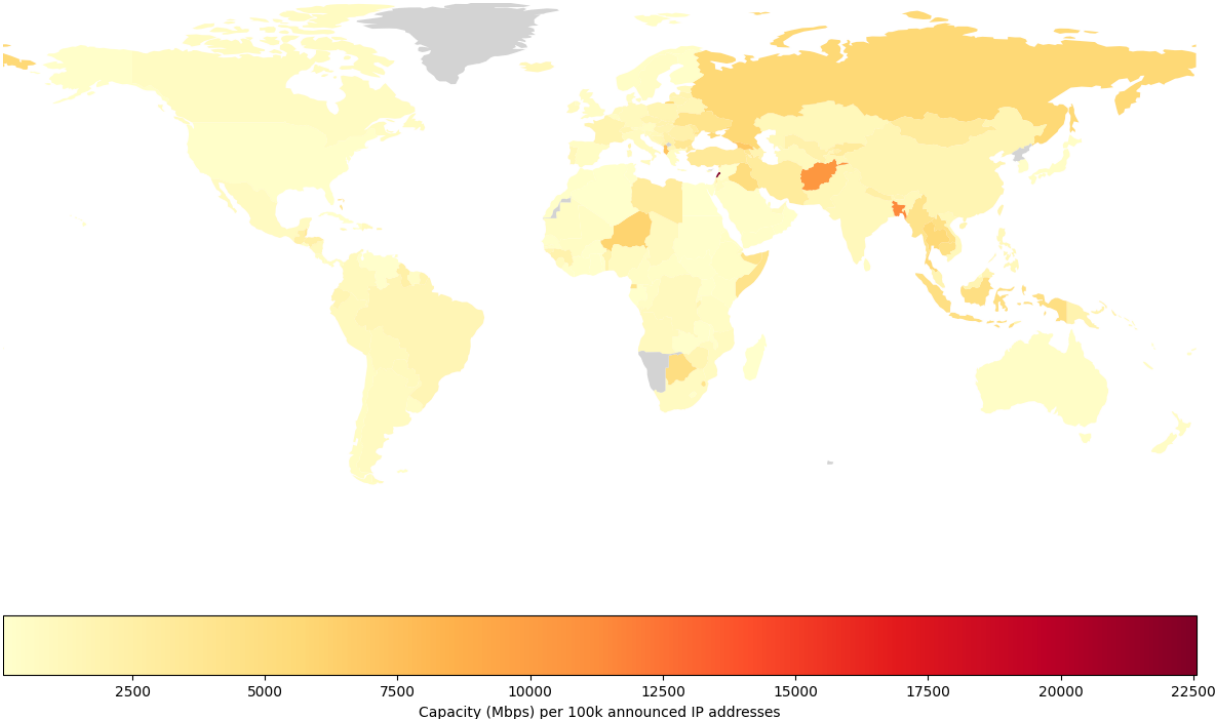

Figure 18: Measured capacity relative to announced IP addresses in May 2020