A Third of the Internet is Under Attack: a Macroscopic Characterization of the DoS Ecosystem

Mattijs Jonker
University of Twente
m.jonker@utwente.nl

Alistair King
CAIDA, UC San Diego

Johannes Krupp
CISPA, Saarland University

Christian Rossow
CISPA, Saarland University

Anna Sperotto
University of Twente

Alberto Dainotti
CAIDA, UC San Diego

ABSTRACT

Denial-of-Service attacks have rapidly increased in terms of frequency and intensity, steadily becoming one of the biggest threats to Internet stability and reliability. However, a rigorous comprehensive characterization of this phenomenon, and of countermeasures to mitigate the associated risks, faces many infrastructure and analytic challenges. We make progress toward this goal, by introducing and applying a new framework to enable a macroscopic characterization of attacks, attack targets, and DDoS Protection Services (DPSs). Our analysis leverages data from four independent global Internet measurement infrastructures over the last two years: backscatter traffic to a large network telescope; logs from amplification honeypots; a DNS measurement platform covering 60% of the current namespace; and a DNS-based data set focusing on DPS adoption. Our results reveal the massive scale of the DoS problem, including an eye-opening statistic that one-third of all `/24` networks recently estimated to be active on the Internet have suffered at least one DoS attack over the last two years. We also discovered that often targets are simultaneously hit by different types of attacks. In our data, Web servers were the most prominent attack target; an average of 3% of the Web sites in `.com`, `.net`, and `.org` were involved with attacks, daily. Finally, we shed light on factors influencing migration to a DPS.

CCS CONCEPTS

• **Networks** → **Denial-of-service attacks**; Network management;
• **Security and privacy** → *Security services*; Web protocol security;

KEYWORDS

DDoS; reflection attacks; spoofed attacks; cloud-based mitigation

ACM Reference Format:
Mattijs Jonker, Alistair King, Johannes Krupp, Christian Rossow, Anna Sperotto, and Alberto Dainotti. 2017. A Third of the Internet is Under Attack: a Macroscopic Characterization of the DoS Ecosystem. In *Proceedings of IMC '17, London, UK, November 1–3, 2017,* 14 pages.
https://doi.org/10.1145/3131365.3131383

IMC '17, November 1–3, 2017, London, UK

ACM ISBN 978-1-4503-5118-8/17/11…$15.00
https://doi.org/10.1145/3131365.3131383

1 INTRODUCTION

Denial-of-Service (DoS) attacks have rapidly increased in frequency and intensity, with recent reports of attacks reaching *1* Tbps [1]. The rise of the DoS-as-a-Service phenomenon (e.g., booters) [2], has dramatically expanded the population of potential perpetrators, who can now purchase the execution of attacks powerful enough to saturate *1-10* Gbps links. Events like the recent attack against Dyn [3], or the DNS root server system [4], have demonstrated the vulnerability of critical Internet infrastructure to DoS attacks.

The rise of DoS attacks has stimulated a new market for DDoS Protection Services (DPSs), i.e., external services aiming at filtering and dropping malicious traffic before it reaches the intended target. Several authors of this paper have empirically shown an increasing trend in the adoption of DPSs [5]. But a rigorous characterization of the DoS phenomenon itself faces tremendous challenges, rooted in the need for sustained operational infrastructure to capture indicators of a variety of different types of DoS attacks, as well as complex data fusion techniques that must integrate heterogeneous raw data sources as well as meta-data to support classification and correlation of attack events.

We offer a set of contributions toward this goal, by introducing and applying a new framework to enable a macroscopic characterization of attacks, attack targets, and mitigation behaviors. We leverage four distinct data sets that cover a recent two-year period (March 2015 - Feb 2017). We use two raw data sources that provide signals of DoS attack events and complement each other: (1) the UCSD Network Telescope [6], which captures evidence of DoS attacks that involve randomly and uniformly spoofed IP addresses; and (2) the AmpPot DDoS honeypots [7], which witness reflection and amplification DoS attacks – an attack type that involves specifically spoofed IP addresses. Our data sets reveal more than *20* M DoS attacks targeting about *2.2* M `/24` IPv4 network blocks, which is more than one-third of those estimated to be active on the Internet [8, 9]. Furthermore, we discover *137* k cases where both randomly spoofed attacks and reflection and amplification attacks were simultaneously launched against the same target.

We also find that most DoS attacks (e.g., about *69%* for TCP-based attacks) targeted Web servers, so we analyze this prominent class of target in more detail using the OpenINTEL DNS measurement platform. We find that two-thirds of all registered Web domains that we observe were hosted on IP addresses targeted by attacks during our two-year measurement period. On average, on a single day, about *3%* of all Web sites were involved in attacks (i.e., by being hosted on targeted IP addresses). This includes attacks on several large Web hosting companies.

Finally, we study the extent to which such attacks forced Web hosters to migrate to DDoS protection services. Based on OpenINTEL data that specifically focuses on DPS providers [5], we discover that *4.3%* of the attack targets we observe migrate to a DPS following an attack. To understand determining factors that motivate or accelerate migration, we correlate attack duration, repetition and intensity with migration events. While repetition and duration do not significantly influence DPS migrations, we observe that intense attacks significantly accelerate the migration process.

The remainder of this paper is organized as follows. Section 2 provides background on DoS attacks and DPSs. Section 3 describes the four data sets we use. Sections 4, 5, and 6 analyze our comprehensive set of attack events, their impact on Web servers, and the effect of attacks on migration to a DPS, respectively. Section 7 describes related work. Section 8 offers a set of future directions. Section 9 concludes the paper.

2 BACKGROUND

2.1 Denial-of-Service Attacks

DoS is commonly achieved through resource exhaustion, either at the server side (e.g., by sending more requests than it can handle) or at the infrastructure level (e.g., by saturaring a network link). Depending on how attack traffic is generated, DoS attacks can be distinguished into *direct* and *reflection* attacks. *Direct* attacks involve traffic sent directly to the target from some infrastructure controlled by the attackers, e.g., their own machines, a set of servers, or even a botnet under their command. To conceal this infrastructure and to impede countermeasures and attribution, these attacks oftentimes employ random spoofing, i.e., faking the source IP addresses in attack traffic. In contrast, in *reflection* attacks, third party servers are involuntarily used to reflect attack traffic towards the victim. This is possible as connection-less protocols have no means of checking whether a request was sent legitimately or with a (specifically) spoofed IP address. An attacker can thus simply spoof requests in the name of the victim, causing the reflectors' replies to be sent to the victim. To make matters worse, many protocols that allow for reflection also add amplification, causing the amount of reflected traffic sent towards the victim to be many times greater than that sent towards the reflector initially [10] – a problem affecting both old protocols as NTP and IGMP [11, 12] as well as newer protocols such as DNSSEC [13].

Since these attacks try to overwhelm a service by a sheer mass of requests, they are referred to as *volumetric* attacks. Beyond that there are also *semantic* attacks, which do not necessarily aim for resource exhaustion but rather exploit flaws in the attacked services themselves, e.g., by sending a malformed request that causes the service to crash. However, this type of attack has to be tailored specifically to work against a given service, whereas volumetric attacks are service agnostic. In this paper we focus on volumetric attacks.

2.2 DDoS Protection Services

DDoS Protection Services offer means for attack mitigation. They may offer various types of mitigation solutions, which can rely on in-line appliances, require network traffic diversion to the cloud (i.e., the DPS infrastructure), or be a hybrid and do both. Volumetric attacks are typically better dealt with in the cloud, whereas semantic attacks can be mitigated in-line [14, 15]. In this paper, we focus on protection where network traffic diversion is required (i.e., all but strictly in-line solutions). Diversion is usually implemented through the DNS or through the Border Gateway Protocol (BGP).

The DNS can be leveraged for network traffic diversion in a manner similar to how content delivery networks implement load balancing [16]. It is common for DPS providers to combine this approach with a reverse proxy that sits between potentially malicious requests and protected Web sites, so that only benign requests are forwarded to the customer's Web server. Alternatively, the DPS can announce a customer-used BGP prefix (e.g., a /24) to divert all customer-destined traffic to the DPS. Traffic is then scrubbed by the DPS before being sent back to the customer's network by using, e.g., a Generic Routing Encapsulation (GRE) tunnel.

The type of customer and the type of attack determine the potential use of either DNS or BGP. While a hoster with a significant number of Web sites and machines may require BGP-based protection of their entire infrastructure, a DPS customer who needs only to divert traffic destined to a single host (or even a single Web site hosted on a shared server) can do so by relying on the DNS. Our methodology identifies both types of network traffic diversion.

3 DATA SETS

In this paper, we analyze and correlate four data sets, all of which cover a two-year period, from March 1, 2015 to February 28, 2017. The first two data sets contain DoS attack events with different characteristics. Specifically, one contains attack events inferred from backscatter to a large network telescope (Section 3.1.1). The other contains events logged in globally placed amplification honeypots (Section 3.1.2). The third data set is derived by a large-scale, active DNS measurement that provides, among other information, mapping of domain names to IP addresses (Section 3.2). The fourth and final data set tracks which Web sites outsource protection to a DPS (Section 3.3).

3.1 DoS Attack Events

3.1.1 Randomly Spoofed Attacks. The first data set contains attack events inferred from backscatter packets reaching the UCSD Network Telescope [6], a largely-unused /8 network operated by the University of California San Diego. Network telescopes, also called darknets, passively collect unsolicited traffic – resulting from scans, misconfigurations, bugs, and backscatter from denial-of-service attacks, etc. – sent to routed regions of the address space that do not contain any hosts. The UCSD Network Telescope covers approximately *1/256* of the IPv4 address space. Any sizable attack, i.e., one that involves many randomly and uniformly spoofed IP addresses, should therefore be visible on this darknet.

To identify randomly spoofed denial-of-service attacks in the data collected at the telescope, we implemented the detection and classification methodology described by Moore et al. [17] as a Corsaro [18] plugin that we have also released publicly as open source [19]. Our plugin uses the same three-step processes described by Moore et al.: first, we identify and extract backscatter packets, then we combine related packets into attack "flows" based

start	end	#days	source	#events	#targets	#/24s	#/16s	#ASNs
2015-03-01	2017-02-28	731	Network Telescope	12.47 M	2.45 M	0.77 M	31057	25990
			Amplification Honeypot	8.43 M	4.18 M	1.72 M	41678	24432
			Combined	20.90 M	6.34 M	2.19 M	43041	32580

Table 1: DoS attack events data. We consider two years of data from the UCSD Network Telescope and from a DoS amplification honeypot to infer DoS attack events. Over the two years we observe more than *20* million events targeted at more than *2* million /24 network blocks.

on the victim IP address, and finally we perform attack classification and filtering.

Specifically, we classify a packet as backscatter if it is a response packet, i.e., TCP SYN/ACK, TCP RST, ICMP Echo Reply, ICMP Destination Unreachable, ICMP Source Quench, ICMP Redirect, ICMP Time Exceeded, ICMP Parameter Problem, ICMP Timestamp Reply, ICMP Information Reply, or ICMP Address Mask Reply. We then aggregate such packets into flows based on the victim IP address (i.e., the source IP address of the backscatter packets), and we expire flows using the same conservative *300* second timeout described by Moore et al. In the final attack classification and filtering step, we compute statistics about the number of unique spoofed source IP addresses, the number of different ports used, and four metrics of estimated attack intensity: the overall number of packets and bytes, the attack duration, and the maximum packet rate per second (in any given minute). We use the same conservative thresholds described by Moore et al. to filter low-intensity attacks, discarding those with: (i) fewer than *25* packets, (ii) a duration shorter than *60* seconds, and (iii) a maximum packet rate lower than *0.5* pps.[1] While the maximum packet-rate can be used as an indicator of the attack intensity, this statistic also reflects the capability of the victim to endure the attack. That is, a high-intensity attack to a well-provisioned victim will likely result in a higher observed maximum packet rate than the same attack directed at a poorly-provisioned victim.

3.1.2 Reflection and Amplification Attacks. The second data set contains events logged by AmpPot [7]. This honeypot aims to track reflection and amplification DoS attacks by mimicking reflectors. To be appealing to attackers, AmpPot emulates several protocols known to be abused.[2] This way, AmpPot can be found by attackers scanning for reflectors and be "abused" in subsequent DoS attacks.

During an attack, an attacker sends spoofed requests allegedly coming from the victim to AmpPot. In order not to cause harm in actual attacks, AmpPot only replies to sources sending less than three packets per minute. However, recording these requests allows us to infer various information about the attack, including the IP address of the victim, the start and end of the attack, but also the request rate, which can be used as a measure of intensity. To distinguish attacks from other traffic (e.g., scans for reflectors), we only consider events exceeding *100* requests.

An initial set of eight honeypots was installed in November 2014. The set has since been expanded to *24* honeypots. To prevent skew in the dataset by either country or autonomous system, the honeypots are distributed both geographically[3], as well as logically, among various cloud providers and machines operated by volunteers. It has been shown that by making the honeypots attractive to attackers (in terms of the the amplification that attackers can achieve), *24* honeypot instances are sufficient to catch most reflection and amplification DoS attacks on the Internet [7].

3.1.3 Attack Coverage and Target Metadata. Many types of DoS attacks involve spoofed IP addresses. Any sizable DoS attack that involves randomly and uniformly spoofed IP addresses should be visible on the UCSD Network Telescope. Moreover, *24* honeypot instances catch most reflection and amplification attacks, which involve specifically spoofed IP addresses (i.e., that of the victim). Our data sets of attack events therefore complement each other in terms of the DoS attack types that they register.[4]

Table 1 summarizes both data sets. The telescope data set has *12.47* M randomly spoofed attack events, involving *2.45* M unique targets (i.e., unique IP addresses). The honeypots data set has *8.43* M reflection attacks, targeting *4.18* M unique targets. We defer a further discussion of these data sets until Section 4. Both data sets of attack events contain target IP addresses to which we add metadata on geolocation using *NetAcuity Edge Premium Edition* data [20]. We also add metadata on BGP routing by using *Routeviews Prefix-to-AS mappings* data [21].

3.2 Active DNS measurements

The telescope and honeypots data sets contain per attack event the IP address of the attacked target. To evaluate the potential effect of attacks on the Web we need a historical mapping between Web sites and the IP addresses on which they were hosted. To obtain this mapping, we rely on the large scale, active DNS measurement performed by the OpenINTEL platform[5] [22]. The OpenINTEL platform collects daily snapshots of the content of the DNS by structurally querying all the domain names in a full zone, i.e., Top-Level Domain (TLD), for their Resource Records (RRs). The measurement data includes IP address mappings, i.e., `A` records. In this study we identify the Web sites that are potentially affected by attacks by looking for `A` records on `www` labels that, at the time of an attack, resolved to the attacked IP addresses.[6]

We use a subset of the TLDs that OpenINTEL measures. Table 2 shows the details of this data set. We use data for the three generic TLDs (gTLDs) `.com`, `.net`, and `.org`. For each of the three gTLDs,

[1] A packet rate of *0.5* pps to the telescope corresponds to an estimated packet rate of *128* packets per second to the victim (the number should be multiplied by *256*).

[2] The protocols QOTD, CharGen, DNS, NTP, SSDP, MSSQL, RIPv1, and TFTP.

[3] *11* honeypots are located in America, *8* in Europe, *4* in Asia and *1* in Australia.

[4] Attacks in which network traffic is sent to victims directly (e.g., by botnets that do not spoof source IP addresses) are not covered by the two data sets that we use.

[5] https://openintel.nl/

[6] The presence of a `www` label in the DNS is taken as an indicator that Web content was present (or intended) at the time of an attack. We did not probe for Web content.

we show the total number of Web sites over the two-year period. For example, for .com (the largest TLD), a total of *173.7* million Web sites were seen. The *data points* column shows the total number of collected data points, examples of which are CNAME and A RRs. The total number of data points is *1.258* trillion. The *size* column shows the size of the compressed measurement data using Apache Parquet columnar storage [23], with a total of *28.4* TiB. The three gTLDs cover roughly *50%* of the global domain namespace [24]. On the last day of the studied, two-year period the three gTLDs account for *153* million active www domain names.

start	#days	source	#Web sites	#data points	size
2015-03	731	.com	173.7 M	1045.9 G	23.5 TiB
		.net	21.6 M	121.0 G	2.8 TiB
		.org	14.7 M	90.7 G	2.1 TiB
		Combined	210.0 M	1257.6 G	28.4 TiB

Table 2: Active DNS data set. We use two years of DNS data collected by the OpenINTEL platform to infer Web sites and associated IP addresses for the .com, .net, and .org gTLDs. In this data set we find *210* M domains that we classify as Web sites (i.e., those with a www label).

3.3 DDoS Protection Services

We are interested in understanding if DoS attacks prompt Web sites to oursource protection to a DPS. Our data set on DPS providers contains usage information for all Web sites in the three previously mentioned gTLDs. We created this data set by using the methodology that we previously published in [5]. This methodology relies (also) on OpenINTEL data. The created data set covers the use of ten DPS providers. Nine out of ten are leading commercial providers [25]. Specifically, these are Akamai, CenturyLink, CloudFlare, DOSarrest, F5 Networks, Incapsula, Level3, Neustar, and Verisign (as in [5]). The tenth is an extension; we added VirtualRoad, a non-commercial provider that protects Web sites run by journalists, activists, and human right workers. By adding VirtualRoad we include in our analysis also attack targets that would not normally outsource protection to commercial DPS. Table 3 shows the details of the data set in terms of the total number of Web sites that we associate with each of the ten providers, over the two-year period. In Section 6, by correlating the DPS use data set with the DoS attack events and the active DNS measurements data sets, we study if (and when) Web sites start outsourcing protection following an attack.

4 ANALYSIS OF ATTACK EVENTS

A third of the Internet is under attack. Together, our data sets of attack events account for *20.90* M attacks, targeting *6.34* M unique IP addresses, over a two-year period (Table 1). We observe a total of *2.19* M unique /24 network blocks that host at least one target, which is about a third of the ~*6.5* M /24 blocks recently estimated to be active on the Internet [8, 9]. For repeated attacks against the same IP address, we see fewer events per target IP in the honeypots data than in the telescope data, which we attribute to more follow-up in randomly spoofed attacks. Combined numbers for both data sets also show overlap in targets, which we investigate further in this section.

provider	#Web sites
Akamai	5.86 M
CenturyLink	0.87 M
CloudFlare	4.27 M
DOSarrest	7.04 M
F5	3.58 M
Incapsula	3.78 M
Level 3	0.47 M
Neustar	10.78 M
Verisign	4.34 M
VirtualRoad	< 100

Table 3: DDoS Protection Service use. For each of the *10* DPS providers that we consider, we identify the Web sites they provide protection services for by using the DNS data from OpenINTEL.

Around *30* k DoS attacks a day are visible. Figure 1 shows statistics over time for the two years' worth of attack events. The top graph shows randomly spoofed attacks, i.e., those in the telescope data set. The *attacks* curve shows the number of events seen on each day, which averages out to about *17.1* k daily. The *unique targets* curve is noticeably lower than the *attacks* curve, in each day, highlighting that some targets are hit more than once on the same day by randomly spoofed attacks.

The middle graph of Figure 1 shows statistics over time for attack events in the honeypots data set. The average number of *attacks* is about *11.6* k daily. In this case, the *unique targets* and *attacks* curves are not as far apart as for randomly spoofed attacks, reflecting a lower average number of events per target IP address.

Finally, the bottom graph in the same figure shows the combination of attack events from both data sets. In total, we observe an average of *28.7* k attacks per day. The curve of *unique targets* is not the sum of the unique targets seen in each data set individually. This is because some targets are hit by both randomly spoofed and reflection DoS attacks on the same day, which we investigate in more depth at the end of this section.

The combined events as well as the individual time series reveal spikes and plateaus in terms of the number of attack events. We evaluate outliers in Section 5, where we study the potential effect of (intense) attack events on the Web. A takeaway from these results is that each day we see attacks on tens of thousands of unique target IP addresses, spread over thousands of autonomous systems, as shown by the *targeted ASNs* curves.

By-country target ranking follows Internet space usage patterns, with some notable exceptions. We rank the most-commonly targeted countries, based on the geolocation metadata of target IP addresses. Table 4a shows that more than one fourth of randomly spoofed attack targets geolocate to the United States, with *25.56%* (or *625* k) of all unique IP addresses. China follows second, with *10.47%* of targets. These two countries also rank first and second for reflection attacks in Table 4b, respectively with *29.5%* and *9.96%* of *4.18* M unique target IP addresses. In general, we find that the

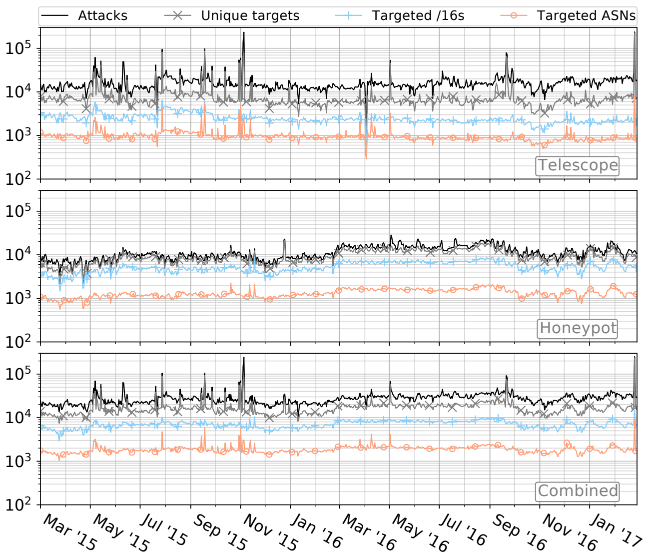

Figure 1: The number of attacks over time (black lines), and the number of IP addresses (grey lines), /16 network blocks (blue lines), and ASNs (orange lines) targeted over time for: randomly-spoofed DoS attacks observed in the telescope data set (top graph), attack events in the honeypots data set (middle graph), and the union of these two data sets (bottom graph). Note that the combined data is not simply the sum of the top two graphs: in some cases we observe targets attacked by both randomly-spoofed, and reflected DoS attacks, on the same day.

country	#targets	%
US	625 k	25.56%
China	256 k	10.47%
Russia	140 k	5.72%
France	126 k	5.14%
Germany	103 k	4.20%
Other	1200 k	48.91%

(a) Telescope

country	#targets	%
US	1232 k	29.50%
China	416 k	9.96%
France	323 k	7.73%
GB	266 k	6.37%
Germany	216 k	5.18%
Other	1727 k	41.26%

(b) Honeypot

Table 4: The targeted IP addresses and percentage of all observed attacks per-country (based on the NetAcuity Edge IP geolocation database). While this ranking mostly follows Internet space usage patterns, we find some notable exceptions, e.g., while Japan ranks *3rd* in recent address space usage estimates, it ranks *25th* and *14th* in the telescope and honeypots data respectively. On the other hand, Russia and France rank higher in terms of attacks compared to address space usage.

IP protocol	TCP	UDP	ICMP	Other
events (%)	79.4%	15.9%	4.5%	0.2%

Table 5: IP protocol distribution. The percentage of all attacks per IP protocol as observed in the telescope data.

type	#events	%
NTP	3.38 M	40.08%
DNS	2.21 M	26.17%
CharGen	1.89 M	22.37%
SSDP	0.71 M	8.38%
RIPv1	0.23 M	2.27%
Other	0.01 M	0.73%

Table 6: Reflection protocol distribution. Number of attacks (and percentage of all attacks) per reflection protocol as observed in the honeypots data.

two rankings are largely consistent and mostly reflect available statistics of Internet address space utilization (e.g., routed space or estimated used space [26]). However, there are some notable exceptions. While in recent estimates Japan ranks third (*6.22%* and *6.33%* of space announced on BGP or inferred as actively used, respectively [27]), in the telescope and honeypots data sets it ranks *25th* and *14th*, respectively. Russia and France, are instead examples of countries that in these attack datasets rank higher than in estimates of Internet space usage. In the case of France, we found out that this shift is mostly due to attacks to OVH, a large hoster that was heavily attacked in 2016 [1].

TCP is the preferred protocol in randomly spoofed attacks. The distribution of IP protocols in the attack events in the telescope data set provides an overview of the flooding approach used. Table 5 shows that the majority of these attacks involve TCP (*79.4%*), while UDP and ICMP follow at *15.9%* and *4.5%*, respectively. ICMP in this distribution denotes ICMP attack traffic (e.g., a ping flood, which leads to echo reply backscatter). In case an ICMP unreachable message reaches the telescope, we register the protocol of the quoted packet, e.g., UDP for a UDP packet that could not reach its destination. Other protocols (e.g., IGMP) account for *0.2%* of attack events.

NTP is the preferred reflector protocol in reflection and amplification attacks. The honeypots data set does not suggest which specific service was targeted by reflection attacks. Instead, we observe which amplification vector (i.e., reflector protocol) was used by the attacker. Table 6 shows a distribution of the protocols chosen by attackers. NTP leads with *3.38* M attack events, accounting for *40.08%* of the *8.43* M reflection attacks seen over two years (Table 1). The second and third placed, DNS and CharGen, account for *26.17%* and *22.37%*, respectively. Examples of protocols following SSDP and RIPv1 in terms of occurrence are MS SQL and TFTP.

NTP is also the most-used protocol for reflection according to various vendor reports. While we find similarities between our results and vendor reports, we also find differences. As vendor reports are based on customer-specific data and oftentimes do not state the scientific method used, we do not delve into these similarities and differences further.

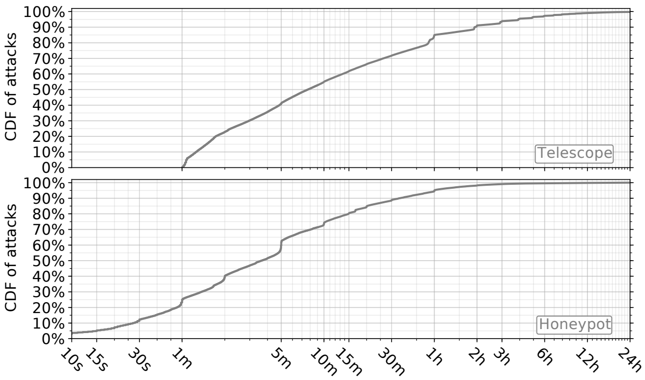

Figure 2: Duration of attacks. The distributions of duration in the telescope (top graph) and honeypots (bottom graph) data sets.

Randomly spoofed attacks tend to last longer. 10% last more than an hour and a half. Each target is attacked a certain amount of time. Attacks typically last minutes up to hours. Figure 2 shows the distributions of the attack duration in our data sets. The top and bottom graphs refer to randomly spoofed and reflection attacks, respectively. About *40%* of randomly spoofed attacks last five minutes or shorter. Attacks in the telescope data set last at least one minute due to the minimum duration threshold that we outlined in Section 3.1.1. We find that roughly the top *10%* of randomly spoofed attacks last *1.5* hours or longer. While attacks in the telescope data set can last longer than a day, these cases are rather scarce (∼*0.2%*). The mean duration is *48* minutes and the median is *454* seconds.

For attack events in the honeypots data set we find that *50%* of attacks last *255* seconds or shorter. The top *10%* of attacks last *40* minutes or longer, and roughly *6%* of attacks last an hour or longer. The mean attack duration is *18* minutes and the median duration is *255* seconds. We note that because of how the honeypots operate, they cap attack event durations at *24* hours. As only ∼*0.02%* of attacks last *24* hours we don't expect this cap to significantly affect the results.

More than a thousand attacks of medium to maximum intensity occur on a daily basis. The attack data sets contain intensity attributes, which we use to analyze the strength of attacks. For randomly spoofed attacks we see the maximum number of packets per second reaching the telescope during the attack. This rate can range from tens to tens of millions of packets per second. To infer an estimate of the packet rate reaching the victim, assuming the attack is using uniformly random spoofing, the rate should still be multiplied by *256* (Section 3.1.1). For reflection attacks we observe the average number of requests made to the reflector per second. This number can range from below one to hundreds of thousands. The reason for the comparative difference in the higher ranges is because reflection attacks are amplified, and need fewer packets to reach large traffic volumes.

We use these attributes to estimate the intensity distributions over attacks. Figure 3 shows the results for attack events in the telescope data set. A steep curve shows that about *70%* of attacks generate only about two backscatter packets per second (max) reaching the telescope, which translates to an estimated attack rate of *512* packets per second to the victim. For about *17%* of the attacks, the telescope observes more than *10* packets per second (an estimated attack rate of *2560* packets per second to the victim). The mean and median values are *107* and *1*, respectively.

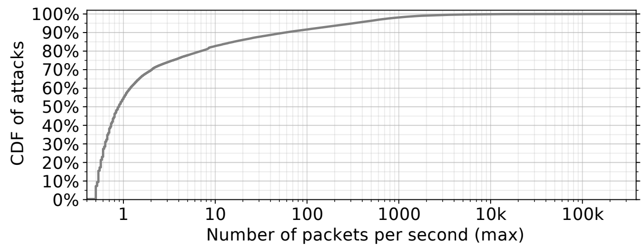

Figure 3: The intensity distribution for attack events in the telescope data set. The number of packets per second (max) should be multiplied by *256* to estimate the the packet rate reaching the victim.

As a honeypot is part of a larger group of amplifiers used during an attack, the attack intensity depends on the total number of amplifiers involved. While it is unclear how many other amplifiers are involved in each attack, our best guess is that the total number of amplifiers will not vary significantly among attacks using the same amplification vector. Therefore, in Figure 4 we show the intensity distribution separately per protocol for the honeypots data set. We show the overall distribution for all attack events, as well as separate curves for the top five used reflector protocols.[7] For most protocols, about *70-90%* of attacks see a gradual increase in the number of requests per second, starting as low as below one on average, to a couple thousand. The number of requests involved clearly varies per protocol. Taking NTP as an example, roughly the first *90%* of attacks see up to *2000* packets per second, whereas the top intensities involve tens to hundreds of thousands of packets per second. These distributions are also different compared to the telescope data, which we attribute to the different nature of attack events. The overall mean and median values are *413* and *77* requests per second, respectively.

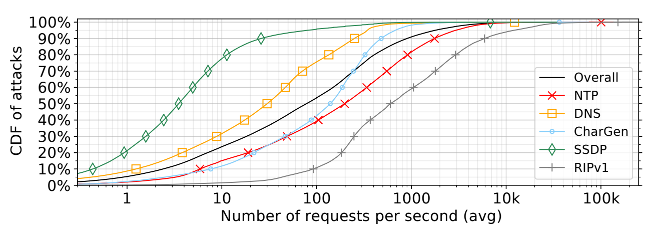

Figure 4: The intensity distribution for attack events in the honeypots data set. We show the distribution for the top five reflector protocols used, as well as the overall distribution.

[7]Note that these five reflector types are involved in all but *10*k attack events, as shown in Table 6.

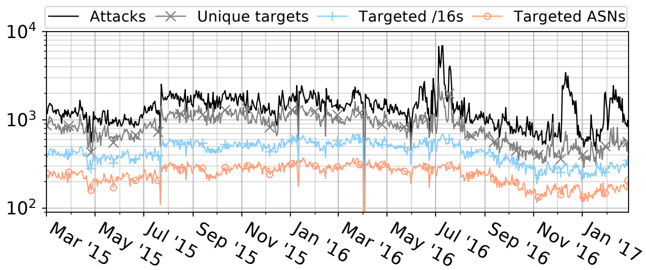

Figure 5: High-intensity attack events over time. The number of attacks with a medium or higher intensity, per day, for the telescope and honeypots data sets combined.

type	#events	%
single-port	7.56 M	60.6%
multi-port	4.91 M	39.4%

Table 7: Number of target ports distribution. Number of attacks (and percentage of all attacks) per target port cardinality in the telescope data.

Figure 5 shows attack events that have a medium intensity or higher, over time, for both data sets combined. We consider an attack event to be of medium intensity or higher if its intensity is at least the mean of all intensities in the corresponding data set. On average, daily, we observe *1.4* k attacks within this intensity range, compared to the overall average of *28.7* k attacks per day (Figure 1). We study one of the peaks visible in the curve in Section 5.

Web, gaming, and MySQL ports are the most attacked in randomly spoofed attacks. Randomly spoofed traffic sent to flood a victim can target one or multiple ports. One reason to target a single port is because the attacker wants to take down a specific networked daemon. Another reason is because the port is known (or assumed) not to be filtered by a firewall. Table 7 shows, for the *12.47* M attack events of this type, the number of events that targeted strictly one port (*60.6%*), as well as those that involved multiple ports (*39.4%*).[8]

We map the ports of attacks that target only a single port to applications, i.e., services, on the basis of both IANA port assignments, as well as commonly used port numbers. Table 8 shows the results of this mapping for TCP and UDP. We show in Figure 8 per protocol the top five potentially targeted services, along with their share of the distribution within that respective protocol.[9]

Table 8a shows the results for TCP. HTTP ranks first with *2.83* M attack events, which account for *48.68%* of *5.81* M single target port attacks on TCP. HTTPS ranks second with *20.68%*. The third place goes to MySQL (3306/TCP), with a share of *1.12%*, which is significantly lower than HTTP(S). For UDP, in Table 8b, the most-attacked port is associated with various on-line multiplayer games

[8]For the honeypots data we do not make a port number distinction, because we do not keep track of the typically ephemeral target port in the reflected packet.

[9]We say "potentially" because we do not know if the service was listening at the time of the attack. Moreover, port might have been chosen by an attacker merely to penetrate a firewall to perform a service agnostic attack.

type	#events	%
HTTP	2.83 M	48.68%
HTTPS	1.20 M	20.68%
MySQL	0.06 M	1.12%
DNS	0.06 M	1.07%
VPN PPTP	0.06 M	0.99%
Other	1.60 M	27.46%

(a) TCP

type	#events	%
27015	225.4 k	18.54%
37547	24.8 k	2.04%
32124	17.1 k	1.41%
28183	16.9 k	1.39%
MySQL	15.8 k	1.30%
Other	916 k	75.32%

(b) UDP

Table 8: The distribution of target ports in the telescope data set. We show the top five potentially targeted services – based on IANA port assignments – for randomly spoofed attacks to a single port using TCP (left) and UDP (right).

and the Steam platform.[10] About *75%* of the UDP attack events target ports that do not rank among the top five, which is because these attacks are spread out over the roughly *65* k remaining ports.[11]

There are two important takeaways from these results. First, while attacks associated with on-line gaming are most apparent for UDP, most other attack events for UDP are spread out over the full port range. Second, more than two thirds of all attack events over TCP potentially target Web infrastructure (*69.36%*).

Randomly spoofed attacks against Web ports are more intense. Given the prominent presence of Web ports (i.e., 80 & 443) in the telescope data set we evaluate the mean and median intensity of attacks that potentially target Web ports. We find that the mean (maximum per attack) rate observed at the telescope is *226* packets per second – corresponding to an estimate of almost *60* k packets per second. This is a change upward from *107* for all randomly spoofed attacks (the median remains the same). We also compared the duration statistics with their overall counterparts and find that the mean drops to *10* minutes (down from *48*) and the median drops to *240* seconds (down from *454*). We thus find that attack events that involve Web ports are more intense than the overall, while lasting shorter. These attacks might have an adverse effect on Web sites (Section 5) and trigger outsourcing protection to a DPS (Section 6).

Randomly spoofed and reflection and amplification attacks are sometimes used jointly against the same target. Finally, we study cases in which targeted IP addresses show up in both the telescope and the honeypots data sets. That is, the targets are hit by randomly spoofed attacks as well as reflection attacks over time. The telescope and honeypots data sets have *282* k unique target IP addresses in common (Table 1). Out of *282* k targets, *137* k were hit simultaneously by joint attacks, i.e., attacks that overlap in time. An example of a joint attack is a SYN flood combined with an NTP reflection attack. The vast majority (*77.1%*) of randomly spoofed attacks co-participating in attacking a victim involve a single port in the telescope data set: we see an increase from *60.6%* (Table 7), suggesting that joint attacks are more likely to target a specific service. The target port distribution of randomly spoofed attacks jointly involved with reflection attacks has more attacks to *27015/UDP* (*53%* up from *18.54%*), which suggests that joint attacks

[10]http://steampowered.com/

[11]A few examples of services over UDP that follow the fifth placed MySQL are NTP (123/UDP) and NetBIOS (138/UDP).

might be used to gain an edge in on-line gaming. For TCP, an increase in HTTP from *48.68%* to *50.23%* is seen. While the latter is a subtle change, it could indicate that serious attackers, i.e., those who launch both randomly spoofed and reflection attacks, target Web services more often.

The distribution of IP protocols in randomly spoofed joint attacks is similar to that of all randomly spoofed attacks and shifts only by tens of percents. For reflection attacks co-participating in attacking a victim, we find that CharGen's use drops by half, to *11.5%*, while the other four protocols in the top five gain. NTP gains most with an increase to *47.0%*.

The autonomous system most-commonly targeted by joint attacks is *AS12276* (*OVH*), with *12.3%* of *137* k unique joint attack targets. *China Telecom* is placed second with *5.4%*. *China Unicom*'s *AS4837* is third (*3.1%*). When considering joint attacks, the per-country distribution does not differ significantly from those for single attacks.

The first-most and second-most countries to which joint targets geolocate are the US and China, with *24.4%* and *20.4%*, respectively. France comes third (*9.5%*) and Germany fourth (*6.5%*). These four countries are also in both top fives in Table 4a and Table 4b, and in the same order. Russia, which was not in the top five for reflection attacks, is fifth placed for joint attacks (*4.1%*).

5 THE EFFECT OF ATTACKS ON THE WEB

In this section we evaluate the potential effect of attack events on the Web. We consider the subset of attack events that target IP addresses for which we can determine Web site associations, using the active DNS measurement data set described in Section 3.

While analyzing Web site associations we may find that multiple Web sites share an attacked IP address. As a consequence, an attack on a single IP can potentially affect millions of Web sites simultaneously. These cases occur when an IP address is used by a larger party, such as a hoster. In case of multiple associations, a single Web site as well as the hoster as a whole may have been the intended target of the attack. Regardless, all Web sites that share that IP address can potentially be affected. We identify large parties by looking at routing information for the attacked IP address, by looking at a common name server in the `NS` record, or a common `CNAME` through which Web sites expand to the shared IP address. To elaborate the last point: in some cases a `CNAME` record in the DNS can reveal more about a Web site than the Web site's IP address. For example, some hosters rely on Amazon AWS, which means that IP routing information points to Amazon and not to the hoster. A customer-specific `CNAME` that all Web sites share might still reveal the hoster.

Many target IP addresses belong to large hosters, with each mapping up to millions of Web sites. We find Web site associations on *572* k of the *6.34* M unique target IP addresses in the attack events. This means that of uniquely targeted IP addresses, at least *9%* host one or more Web sites. Figure 6 shows the number of Web sites affected by attack events. Each bar, i.e., bin, represents a "co-hosting" group, which indicates how many Web sites were associated with a targeted IP address at the time of an attack. The magnitude of each group is the number of target IP addresses

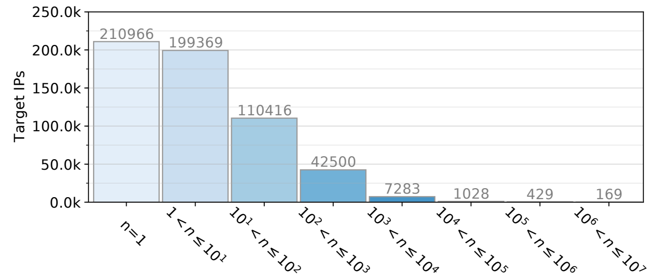

Figure 6: Web site associations with IP addresses targeted by attacks. Each bar indicates the number of unique target IP addresses (y axis) associated with a number of Web sites within a given bin (x axis).

within the group.[12] More than a third of these IP addresses (∼*211* k) were associated with a single Web site at the time of an attack, whereas, at the other end of the distribution, *169* targets hosted *1* M to *3.6* M Web sites potentially affected by the attack event (*3.6* M is the maximum in the right-most group in the graph).[13]

The active DNS data set used in this work covers only `com`, `net`, and `org` Web sites. Our estimate of Web sites per target IP is therefore a lower bound. An IP address in the *n=1* "co-hosting" group might be associated with a Web site in another TLD (e.g., `www.example.tk`). This would knock the target IP out of the group in question, and thus affect the distribution. To analyze this effect we considered `com`, `net`, and `org` individually. The shape of Figure 6 is similar for the three individual distributions, which suggests that the distribution among "co-hosting" groups would not drastically change even if we considered more TLDs.

Isolating Web targets reveals an even more pronounced majority of TCP-based randomly spoofed attacks and NTP-based reflection attacks. Randomly spoofed attacks against IP addresses that host Web sites primarily make use of TCP. Specifically, *93.4%* of attack events, up from *79.4%* for all attacks (Section 4). Moreover, attacks on targets that host Web sites mostly also target Web infrastructure ports: *87.60%*, up from *69.36%* for all attacks. We find that NTP is the most commonly used reflector type on such targets: *54.69%*, up from *40.08%*.

Over two years, 64% of inferred (.com, .net, .org) Web sites were hosted on IP addresses targeted by attacks. On average, 3% of Web sites were involved daily. Figure 7 shows, for every day in our two-year observation period, the total number of Web sites associated with attacked target IP adresses on that day. The top graph is for all attack events, and the bottom one is for attack events with a medium to high intensity. In each graph, the grey curve shows the number of Web sites (potentially) affected, in millions, whereas the black curve shows the (smoothed) percentage

[12]Each IP address can contribute once to a "co-hosting" group in this visualization.
[13]This maximum is found on a target IP that is routed by DOSarrest, one of the DPS providers considered in this work. Google and Amazon are other examples with (various) IP addresses in this group.

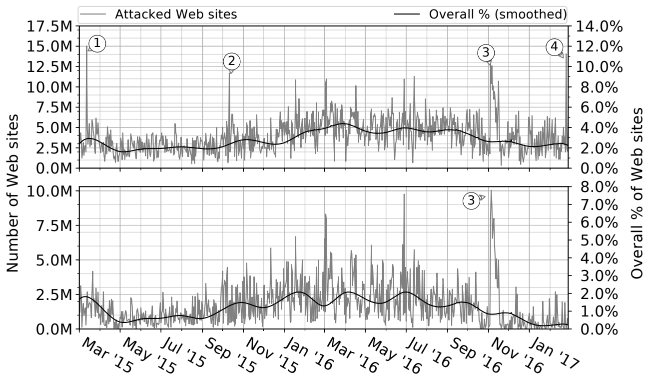

Figure 7: Web site associations with attacked targets over time. The number of Web sites on attacked IP addresses for all attacks (top graph) and medium to high intensity attacks (bottom graph). The left y axis shows the number of Web sites and the right y axis shows the percentage of all Web sites in the measured namespace.

that the involved Web sites make up of all Web sites in the measured namespace (right y axis).[14]

We link almost *134* M unique Web sites to all attack events observed over the two-year period. The average number of Web sites is just under *4* M per day, which translates to about *3%* of all Web sites in the measured namespace.[15] For attacks of medium intensity of higher, the average is *1.7* M daily (*1.3%*).[16] The fraction of Web sites that are potentially affected daily is considerable, which does not come as a surprise given the large number of ASNs and /24 prefixes that we see attacked daily (Section 4).

Investigating attacks to large hosters. In the number of affected Web sites, various peaks are discernible, the largest of which involves *11.82%* of all Web sites. We evaluate this peak, along with three others, as examples of the potential effect of attacks on the Web.

The first peak (cf. ① in Figure 7) on March 12, 2015 involves attacks that associate with a little over *15* million Web sites, which is *11.82%* of all Web sites. We identified several third-parties that offer hosting as large targets on this day. A significant number of Web sites are hosted by *GoDaddy*, through a set of about twenty targeted IP addresses, all routed to their AS. A large part is associated with *WordPress*, primarily through two consecutive IP addresses that belong to *Automaticc Inc.*, the company behind *WordPress*. Another IP address routes to the security infrastructure of CenturyLink, one of the DPS providers considered in this work, which shows that the attack was probably mitigated. Many of the target addresses appear as joint attacks in the honeypot and the telescope data sets, with low to medium intensities.

The second peak (cf. ②) on October 10, 2015, involves *11.7* million Web sites. Among the targets we find several large hosters such as *Squarespace* and *OVH*. Another prominent target is a domain names reseller that is hosted in *Amazon AWS*.[17] The third peak we investigate, occurs on November 4, 2016 (cf. ③). It involves a little over *13* M Web sites. About *10* M of these Web sites are hit by an attack of high intensity, as can be seen in the bottom graph. This number is largely made up by *GoDaddy*-hosted Web sites. A significant number is also associated with *Wix.com*, a Web site development platform.[18] *Squarespace* is among the targets. The final example (cf. ④) is for February 25, 2017. This peak involves *14.1* million Web sites, hosted by various companies, such as *GoDaddy*, *OVH*, *Network Solutions*, and a variety of hosting companies that are subsidiaries of the *Endurance International Group (EIG)*.

Overall, the three most frequently attacked larger parties that we identify over the two-year period are, in order, *GoDaddy*, *Google Cloud*, and *Wix*. Other names include *Squarespace*, *Gandi*, and *OVH*.

While our focus in this section is on attacks that affect Web sites, during our analysis we encountered several IP addresses that can be linked to the mail infrastructure of a large number of domain names. It is not www labels that map to these IP addresses, but rather the mail exchanger records (MX) of domain names. For example, we find that *GoDaddy's* e-mail servers, which are used by tens of millions of domain names, are frequently targeted by DoS attacks. In future work, we plan to investigate the impact of DoS attacks on mail infrastructure and for this purpose we recently instrumented our measurement infrastructure to query for more DNS RRs on the names found in MX records.

6 ATTACK EFFECTS ON DPS MIGRATION

In this section we study whether attacks on Web sites have an effect on migration to a protection service, and to which extent. Web site owners who maintain their own hosting, as well as hosting companies that provide hosting infrastructure on a larger scale, may start outsourcing protection to a DPS after being targeted by a DoS attack. Our data set on DPSs allows us to identify, for all the Web sites we infer in com, net, and org, the day of migration to any of the ten providers we consider in this work (within the two year period of analysis). By combining this information with our data sets of attack events, we analyze if, and when, Web sites migrated to a DPS, following one or more attacks.

We define a classification taxonomy for Web sites according to the tree in Figure 8. The root of the tree represents the overall set of domains (over our two year observation period) that we infer to be Web sites (*210* M). We then split this set into two: those for which we observed attacks (*134* M), and those for which we did not (*76* M). We find that the majority of Web sites (*64%*) were observed to be on attacked IP addresses over the course of two years. Then, for both of these categories (`attack observed` and `no attack observed`), we identify those Web sites that either already use a DPS (`preexisting customers`) – either from the beginning of our data set, or the first time they are found in the DNS – and those that do not (`non-preexisting customers`). We find a much higher percentage of `preexisting customers` in domains for which we observed attacks (*24.9* M, *18.6%*) than for those where

[14] For smoothing we interpolate a cubic spline between the median number of affected Web sites per month.

[15] Multi-day attacks, i.e., those that cross day boundaries, count only towards the day on which the attack was started.

[16] Recall from Section 4 that an attack is considered to be of medium intensity if its intensity is the mean of all intensities in its attack events data set.

[17] This company has its own AWS CNAME, which allowed us to identify it even though the IP belongs to AWS.

[18] Wix hosts in AWS, but uses Incapsula for DDoS mitigation, which is something we previously outlined in [5].

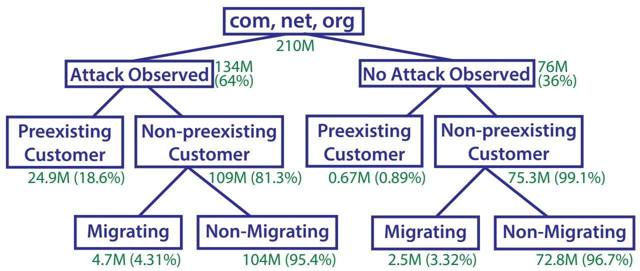

Figure 8: Web site taxonomy. Nodes are annotated with the estimated number of web sites in each category (and the percentage of the parent category population). The root of the tree represents the overall set of domains (over the two years we study) that we infer to be Web sites (i.e., those with a `www` label). We find that of these *210* M Web sites, *64%* were hosted on attacked IP addresses (at the time of an attack) at least once during our two year observation period.

we did not observe attacks (*0.67* M, *0.89%*), suggesting that Web sites we observe to be attacked during our two year observation period may have been previously attacked. Finally, the bottom level of the tree identifies those Web sites that were `non-preexisting customers`, but either did migrate (`migrating`) or did not migrate (`non-migrating`) to using a DPS. In the case of `attack observed` Web sites, we consider a Web site to be `migrating` if it is found in the DPS data set *after* we observed it being under attack. For `no attack observed` Web sites, we consider it to be `migrating` if it is found in the DPS data set *after* it is first seen in the DNS. While we do find a slightly higher percentage of Web sites migrate after an attack (*4.7* M, *4.31%*) compared to those that migrate even when no attack is observed (*2.5* M, *3.32%*), it should be noted that since we do not observe all attacks, the `no attack observed` migrations may still have been influenced by an attack. We also find the percentage of Web sites that either already used a DPS, or during our study migrated to using a DPS, to be much larger for those Web sites that were attacked (*22.1%*) compared to those for which we did not observe an attack (*4.2%*).

While our list of *10* protection services is not exhaustive – AWS (Amazon) and GHS (Google) actually offer DoS protection that we cannot infer and, therefore, the many Web sites they host count towards `non-migrating` in our classification – we take this into account in the following analysis and further discuss this limitation in Section 8. Additionally, because our attack events and DPS datasets cover the same time range, it is possible that we incorrectly classify attacks that occur close to the beginning and/or end of our observation period. More specifically, attacks that overlap the beginning of our DPS data set may have already prompted migration, thus resulting in an incorrect `preexisting customer` classification; similarly, attacks starting near the end may result in migration after our observation period, thus causing in an incorrect `non-migrating` classification. By shortening the observation period of the attacks data by one month on either end and repeating our analyses, we verified that these potential misclassifications have a negligible effect on the overall Web site class distribution.

We manually checked a small sample of Web sites to gain insight into the types of Web sites that are among various combinations of hosting size groups and customer classes.[19] We sampled from the smallest (i.e., $n = 1$), as well as the largest (i.e., $n \geq 10^6$) hosting groups, and for each of the three DPS customer classes (i.e., the leaves of the `attacked` subtree in Figure 8). For the largest hosting group, among those that migrate to a DPS after an attack, we find many Web sites that can be traced to the Wix Web site development platform (also mentioned in Section 5). The Web sites we visited have either personal or business content. Among `non-migrating` within the largest hosting group, we find a lot of landing pages that can be traced to a domain reseller that uses AWS for hosting, as well as personal and business Web sites hosted in Google Cloud. Among the `preexisting customers` we find both personal pages and commercial Web sites such as a Web shop. For the smallest hosting group, we find among `migrating` and `preexisting customers` Web sites that belong to businesses, community Web sites (e.g., related to gaming), and occasionally content for a foundation.[20] For the `non-migrating` class we find, among others, adult Web sites for (video) chat.

Repeated attacks are not a determining factor for migration. We observe a significant fraction (∼*14%*) of Web sites attacked more than once within our observation period. We investigated if the number of attacks experienced by a Web site correlates with migration to a DPS. The top graph in Figure 9 shows the CDF for the distribution of all attacked Web sites as a function of the attack frequency: *7.65%* of these sites are attacked more than *5* times. The bottom graph in the same figure shows instead the CDF, as a function of the attack frequency, for Web sites that migrate to a DPS after an attack event. In this case, the fraction of Web sites that were attacked more than *5* times is *2.17%*. The comparison between the two distributions, suggests that being subject to multiple attacks is not a significant factor in subsequent migration to a DPS.

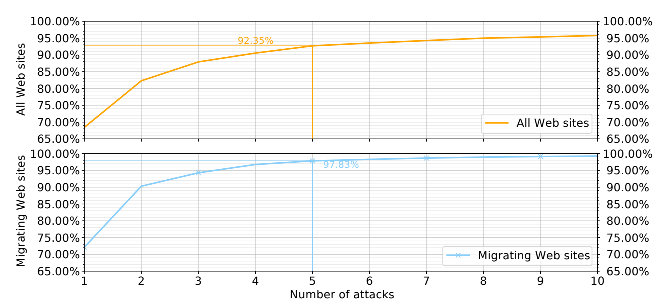

Figure 9: The distributions of attack frequency for all Web sites (top graph) and those that migrate to a DPS following an observed attack event (bottom graph), a comparison of which suggests that being subject to multiple attacks is not a significant factor in subsequent migration to a DPS.

Earlier migration follows attacks of higher intensity. DoS attacks that severely affect Web sites are likely to create an urgency

[19] As outlined in Section 3.2, we did not automatically verify for each potentially affected Web site if content was being served at the time of an attack.

[20] In one case we visited a Web site with radical right content, which may or may not speak to why the Web site was attacked.

Intensity (≤)	0.0	0.07	0.13	0.52	0.85	1.0
Web sites (%)	11.1	95.0	97.5	99.0	99.9	100.0

Table 9: Attack intensity distribution over Web sites. For select percentiles we show the normalized attack intensity in the honeypot and telescope data sets. In case of joint attacks, we take the highest intensity.

to mitigate. This notion makes it reasonable to assume that Web site owners (or hosters) who opt to outsource protection to a DPS will want to do so in an urgent manner. Table 9 shows the normalized attack intensity distribution over attacked Web sites. In the case a Web site is associated with multiple or even simultaneous attacks (e.g., a target IP that appears both in the telescope and honeypots data sets), we pick the highest normalized intensity value.

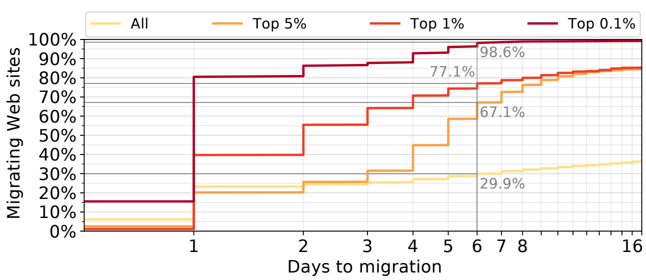

Figure 10: Migration delay for attack intensities. For various percentiles of the normalized attack intensity distribution, ranging from *any* to the *99-th*, we show the number of days it took for Web sites to migrate to a DPS. An urgency to migrate becomes apparent with increasing attack intensity.

Figure 10 shows the cumulative distribution functions of days it took Web sites to migrate, respectively for Web sites attacked with *any* intensity (slowest CDF), and with intensities in the *95-th*, *99-th*, *99.9-th* percentiles of the normalized attack intensity distribution (Table 9). Comparing these CDFs highlights a drastic reduction of the latency between an attack and the effected site migrating to a DPS: almost all (*98.6%*) the top *0.1%* Web sites by attack intensity transition to a DPS within *6* days, whereas for the top *1%*, *5%* and *overall* Web sites only *77.1%*, *67.1%* and *29.9%* of them respectively transition within the same number of days. When considering the Web sites that transition to a DPS within a day from the attack, the difference between the top *0.1%* class and the overall distribution is even more striking: *80.7%* versus *23.2%*, respectively. Differently from the number of attacks, the intensity of a DoS event strongly correlates with migration to a DPS, specifically in terms of speed, which intuitively suggests a sense of urgency in mitigating DoS damage and risks.

Large hosters can potentially skew the mitigation delay distribution by migrating many Web sites at once: if multiple Web sites are associated with an given attack of a given intensity, each Web site counts towards the CDF. We investigated this potential for skew and found that few `migrating Web sites` in the top *97.5-th* percentile were hosted in large numbers.

Attack duration does not strongly correlate with migration. Here we evaluate if attack duration may influence transition to a DPS and specifically timing. As outlined in Section 3.1.1, a target that is brought down by a successful attack will slow down or altogether stop backscattering packets to the telescope. As such, attacks successful enough to trigger migration might be registered with shorter than actual durations in the telescope data set. Amplifiers on the other hand will still receive packets to reflect to the target, and thus have a better sense of the actual attack duration. For these reasons we only consider the durations from the honeypots data set in this analysis.

Overall, we find that the number of days it takes `migrating Web sites` to migrate does not necessarily keep decreasing with an increasing attack duration, unlike is the case for attack intensity. Attacks longer than four hours in duration, which is the top *1%* durations of all honeypot events (Section 3.1.2), lead to the smallest migration delay for `migrating Web sites`. Figure 11 shows the CDF for Web sites affected by attacks within this duration class: of all Web sites associated with attacks that last over four hours, *67.64%* take a day or less to migrate, and *76%* migrate within at most five days. However, more than half of the Web sites that migrate on the next day, following a *4* hours or longer attack, have a common denominator. Specifically, *482* k out of *800* k Web sites trace back to *Wix.com*, who starts outsourcing protection to Incapsula during our observation period. About *18%* take two weeks or longer to migrate, suggesting that duration by itself is not always the deciding factor. We also find common denominators for longer migration delays. For example, *130* k Web sites hosted by *eNom* take more than three months (*101* days) to appear as `migrating Web sites` (of Verisign).

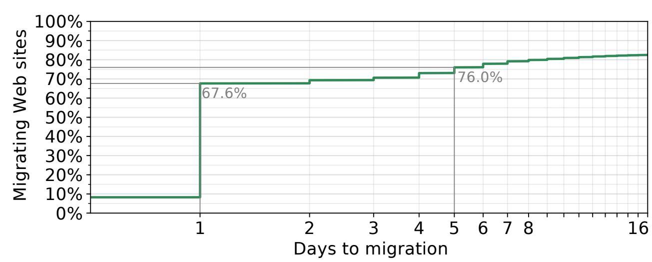

Figure 11: DPS migration delay for longer attacks. The number of days it took for Web sites to migrate to a DPS following attacks with a duration ≥*4* hours.

Finally, we find larger parties that skew the results in favor of, as well as against, short migration delays. Comparing the two previous examples, the first one involves an attack that is three times as intense. Specifically, it involves a normalized attack intensity of *0.18* in the telescope data. This target appeared simultaneously in both the telescope and the honeypots data sets. This finding leads us to conclude that in this case intensity rather than duration was the deciding factor.

7 RELATED WORK

We group related work into three areas of study. The first one pertains to efforts to characterize DoS attacks in general. Such characterizations include, for example, target properties (e.g., geolocation), traffic characteristics (e.g., protocols used), and attacker properties (e.g., malware fingerprinting). The second area is concerned with efforts to measure the effects of attacks. And the third one focuses on attack mitigation.

In 2006, Moore et al. [17] characterized DoS attacks by analyzing events inferred from backscatter packets to a large network telescope. The authors analyze 22 traces of 1-2 weeks each, captured between 2001-2004, totalling *68.7* k events. We incorporated their methodology in our work. Their initial trace is 14 years older than our telescope data set. Comparing results, ours show that the DoS landscape has since changed. As an example, while still dominant, *TCP*'s presence in randomly spoofed attacks has reduced. Moreover, we find a prevalence of single-port attacks.

Krämer et al. [7] and Thomas et al. [28] both present a characterization of attacks from events captured in a set of amplification honeypots. While in both papers the focus is more on reflection attacks in general, in this paper we focus on the correlation with randomly spoofed attacks and on target characteristics. A different view on DoS attacks is given by Santanna et al. [2], who study traffic and source characteristics of the attacks generated on-demand by means of a set of 14 booters. Differently from our paper, this research focuses on the attackers (i.e., the misused infrastructure).

To our knowledge, the last study to characterize DoS attacks at scale by combining multiple, independent data sets dates back to 2006, when Moa et al. used three data sets [29] in their work. Two data sets came from anomaly detection systems and a third was inferred from backscatter. Their analysis covers *35* k attack events, measured over a month, which does not compare in scale with our study. The authors find a *TCP* preference similar to Moore et al., using the same methodology.

More recently, in 2015, Wang et al. [30] analyzed a set of *51* k attack events derived from botnet Command & Control (C&C). Their data set covers a seven-month period and accounts for attacks launched using 674 botnets of 23 different botnet families. They too find joint attacks, in their case by different botnet instances. Furthermore, they show that Web services (i.e., *HTTP*) are the preferred target of many attacks.

The industry regularly releases reports that characterize attacks and trends [31–34]. However, these reports are based on customer-specific data, and oftentimes do not state the scientific method used.

In terms of attack effects, Welzel et al. measured the impact of botnet attacks by monitoring for targets in botnet C&C [35]. Their study covers *646* unique targets, acquired from *14* botnet instances of two botnet families (*DirtJumper* and *Yoddos*). Following attack commands, the authors systematically measure the victims for adverse effects. Occassionally they find that the IP address of a Web site changes following an attack, e.g., in an attempt to mitigate, by pointing it to *localhost*. In a few cases the IP address change is made to (quote) "professional load balancing and DDoS protection services," but this is not investigated further.

Noroozian et al. [36] study the consequences of victimization patterns in targets of DDoS-as-a-Service (e.g., booters). Their focus is on the demographics of the target population. Their results show, among others, that most of the victims are users in access networks, and that the number of attacks in broadband ISP is proportional to the number of ISP subscribers. Similarly to us, their study is also based on two years of data from the AmpPot project. However, we focus in capturing a larger spectrum of attack events by correlating amplification honeypots data with network telescope data.

In terms of effects at a higher level, a DoS attack can have financial consequences for a company, which could face an increase in security costs, or a loss of customers following an attack [37]. While DDoS intensity peaked at *400* Gbps [38] in 2014 and to *600* Gbps in early 2016 [39], the race to the largest DDoS has already reached *1* Tbps in late 2016 with the attack against the hosting company OVH [1]. However, it is not only about how heavy the hammer is, it is also about what it might break. The DDoS attack performed by the Mirai botnet against the service and DNS provider Dyn [3] has provoked a cascading effect that prevented East Coast users to access services such as Twitter, Spotify, or Reddit.

As for mitigation, although the concept of regional cleaning center was already described in 2004 [40], in recent years DDoS protection services have become more and more popular. In previous work we showed a clear trend in adoption [5], but we did not investigate if there is correlation between attack events and migration. To the best of our knowledge no other work addresses the link between attacks and DPS use at scale.

8 FUTURE WORK

We imagine several directions to improve the coverage and depth of our measurement and analysis system:

- We provide a comprehensive view of randomly spoofed and reflection and amplification attacks, but a bigger challenge is development and integration of other attack data sources, e.g., unspoofed volumetric attacks, semantic attacks. By demonstrating the utility of a platform for this type of data fusion, we hope to inspire the community to consider what cooperation would be required to expand the set of data sources.
- Operating such a platform continuously would allow to eliminate the bounding problem, i.e., we do not know which attacks took place before, nor do we know which Web sites migrated to a DPS after, our observation period.
- We examined migration to only ten DPS providers, so we mistakenly infer instances of migration to some other form of protection (e.g., Google) as `non-migrating`. For now we avoid making claims related to the non-migrators, but a more comprehensive view of the DPS ecosystem would improve the fidelity of our inferences.
- We currently consider *50%* of the global DNS name space, a constraint of the OpenINTEL DNS measurement infrastructure. If OpenINTEL could expand to obtain visibility of other Top Level Domains, we would expand our ability to identify and characterize attacks on Web sites.
- We interpret an `A` record for a `www` domain name as an indicator of Web service, though the IP address may host no Web

site. We could add functionality to validate the existence of a Web server before inferring an impact on its reachability. More generally, we could see if the targeted IP addresses run other services.
- We examined the effect of attacks on the migration of Web sites to protection services. As future work, and without adding any other data sources, we could map targeted IP addresses to authoritative name servers, and study the potential effect of attacks on the DNS itself. A potential effect is, e.g., the migration of an authoritative name server to a DPS.

9 CONCLUSIONS

We have established a framework for a more thorough scientific approach to macroscopic characterization of the DoS ecosystem by systematically integrating and correlating large, diverse data sets captured by existing global Internet measurement infrastructure. We integrated data from a large network telescope, honeypots instrumented to observe reflection DoS attacks, and a platform for large-scale active DNS measurements. We augmented these three sources with meta-data such as BGP prefix-to-AS, IP geolocation, and identifiers of DDoS Protection Services and hosting providers. We then developed functionality to extract macroscopic as well as detailed insights about DoS attacks and their impact on Internet infrastructure. Our analysis demonstrates the potential of sustained operation of such infrastructure, and extensions of our analysis approach, in terms of providing situational awareness and informing Internet research, operations and policy communities about a growing threat to Internet stability and reliability. While most of the measurement infrastructure that enables this work already collects data in near-realtime, a significant challenge is enabling near-realtime data fusion, extraction, correlation and visualization to maximize its utility. Our experience in developing this framework, and performing the rigorous characterization of two years of DoS activity, presents a first step in what we hope can become a badly needed source of longitudinal data about the health of what is now our primary communications fabric.

ACKNOWLEDGMENTS

This work is part of the NWO: D3 project, which is funded by the Netherlands Organization for Scientific Research (628.001.018). This research was made possible by OpenINTEL, a joint project of the University of Twente, SURFnet, and SIDN. This work is supported by the EU Commission via the Horizon 2020 project SISSDEN (project ID 700176). This material is based on research sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHS S&T/CSD) via contract number HHSP233201600010C. This material is based on research sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, Homeland Security Advanced Research Projects Agency, Cyber Security Division (DHS S&T/HSARPA/CSD), BAA 11-01-RIKA and Air Force Research Laboratory, Information Directorate under agreement number FA8750-12-2-0326. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. We thank our shepherd and the anonymous reviewers for their valuable feedback.

REFERENCES

[1] Pierluigi Paganini. The hosting provider OVH continues to face massive DDoS attacks launched by a botnet composed at least of 150000 IoT devices. http://securityaffairs.co/wordpress/51726/cyber-crime/ovh-hit-botnet-iot.html, September 2016.
[2] José Jair Santanna, Roland van Rijswijk-Deij, Anna Sperotto, Rick Hofstede, Mark Wierbosch, Lisandro Zambenedetti Granville, and Aiko Pras. Booters - An Analysis of DDoS-as-a-Service Attacks. In *Proceedings of the 14th IFIP/IEEE International Symposium on Integrated Network Management (IM'15)*, 2015.
[3] Scott Hilton. Dyn Analysis Summary Of Friday October 21 Attack. http://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack/, October 2016.
[4] Giovane C.M. Moura, Ricardo de O. Schmidt, John Heidemann, Wouter B. de Vries, Moritz Muller, Lan Wei, and Cristian Hesselman. Anycast vs. DDoS: Evaluating the November 2015 Root DNS Event. In *Proceedings of the 2016 ACM Internet Measurement Conference (IMC'16)*, 2016.
[5] Mattijs Jonker, Anna Sperotto, Roland van Rijswijk-Deij, Ramin Sadre, and Aiko Pras. Measuring the Adoption of DDoS Protection Services. In *Proceedings of the 2016 ACM Internet Measurement Conference (IMC'16)*, pages 279–285, 2016.
[6] UCSD Network Telescope, 2010. http://www.caida.org/projects/network_telescope/.
[7] Lukas Krämer, Johannes Krupp, Daisuke Makita, Tomomi Nishizoe, Takashi Koide, Katsunari Yoshioka, and Christian Rossow. AmpPot: Monitoring and Defending Against Amplification DDoS Attacks. In *International Workshop on Recent Advances in Intrusion Detection (RAID'15)*, pages 615–636, 2015.
[8] Sebastian Zander, Lachlan L.H. Andrew, and Grenville Armitage. Capturing Ghosts: Predicting the Used IPv4 Space by Inferring Unobserved Addresses. In *Proceedings of the 2014 ACM Conference on Internet Measurement Conference (IMC'14)*, 2014.
[9] Philipp Richter, Georgios Smaragdakis, David Plonka, and Arthur Berger. Beyond Counting: New Perspectives on the Active IPv4 Address Space. In *Proceedings of the 2016 ACM Internet Measurement Conference (IMC'16)*, 2016.
[10] Christian Rossow. Amplification Hell: Revisiting Network Protocols for DDoS Abuse. In *NDSS*, 2014.
[11] Jakub Czyz, Michael Kallitsis, Manaf Gharaibeh, Christos Papadopoulos, Michael Bailey, and Manish Karir. Taming the 800 Pound Gorilla: The Rise and Decline of NTP DDoS Attacks. In *Proceedings of the 2014 ACM Internet Measurement Conference (IMC'14)*, pages 435–448, 2014.
[12] Matthew Sargent, John Kristoff, Vern Paxson, and Mark Allman. On the Potential Abuse of IGMP. *ACM SIGCOMM Computer Communication Review*, 47(1), 2017.
[13] Roland van Rijswijk-Deij, Anna Sperotto, and Aiko Pras. DNSSEC and its potential for DDoS attacks: a comprehensive measurement study. In *Proceedings of the 2014 ACM Internet Measurement Conference (IMC'14)*, pages 449–460, 2014.
[14] Jelena Mirkovic, Sven Dietrich, David Dittrich, and Peter Reiher. *Internet Denial of Service: Attack and Defense Mechanisms (Radia Perlman Computer Networking and Security)*. 2004.
[15] Mehmud Abliz. Internet Denial of Service Attacks and Defense Mechanisms. Technical Report TR-11-178, March 2011.
[16] Erik Nygren, Ramesh K. Sitaraman, and Jennifer Sun. The Akamai Network: A Platform for High-performance Internet Applications. *ACM SIGOPS Operating Systems Review*, 44(3):2–19, 2010.
[17] David Moore, Colleen Shannon, Douglas J. Brown, Geoffrey M. Voelker, and Stefan Savage. Inferring Internet Denial-of-service Activity. *ACM Transactions on Computer Systems*, 24(2):115–139, 2006.
[18] Alistair King. Corsaro, 2012. http://www.caida.org/tools/measurement/corsaro/.
[19] Alistair King. Corsaro RS DoS Plugin, 2012. https://www.caida.org/tools/measurement/corsaro/docs/plugins.html#plugins_dos.
[20] Digital Element. Netacuity edge premium edition. http://www.digitalelement.com/solutions/netacuity-edge-premium.
[21] Routeviews Prefix to AS mappings Dataset (pfx2as) for IPv4 and IPv6. http://www.caida.org/data/routing/routeviews-prefix2as.xml.
[22] Roland van Rijswijk-Deij, Mattijs Jonker, Anna Sperotto, and Aiko Pras. A High-Performance, Scalable Infrastructure for Large-Scale Active DNS Measurements. *IEEE Journal on Selected Areas in Communications (JSAC)*, 34(6):1877–1888, 2016.
[23] Apache Parquet, 2014. http://parquet.io/.
[24] The Domain Name Industry Brief. https://www.verisign.com/en_US/innovation/dnib/index.xhtml. Accessed: 2017-05-01.
[25] Rick Holland and Ed Ferrara. The Forrester Wave™: DDoS Services Providers (Q3 2015). Forrester Research, Inc., July 2015.
[26] Alberto Dainotti, Karyn Benson, Alistair King, Bradley Huffaker, Eduard Glatz, Xenofontas Dimitropoulos, Philipp Richter, Alessandro Finamore, and Alex C. Snoeren. Lost in Space: Improving Inference of IPv4 Address Space Utilization. *IEEE Journal on Selected Areas in Communications (JSAC)*, 34(6):1862–1876, 2016.

[27] Lost in Space: Supplemental: Country Inequality (Interactive). http://www.caida.org/publications/papers/2016/lost_in_space/supplemental/country_inequality/.

[28] D. Thomas, R. Clayton, and A. Beresford. 1000 days of UDP amplification DDoS attacks. In *APWG Symposium on Electronic Crime Research (eCrime 2017)*, 2017.

[29] Z. Morley Mao, Vyas Sekar, Oliver Spatscheck, Jacobus van der Merwe, and Rangarajan Vasudevan. Analyzing Large DDoS Attacks Using Multiple Data Sources. In *Proceedings of the 2006 SIGCOMM Workshop on Large-scale Attack Defense (LSAD'06)*, pages 161–168, 2006.

[30] An Wang, Aziz Mohaisen, Wentao Chang, and Songqing Chen. Delving into Internet DDoS Attacks by Botnets: Characterization and Analysis. In *Proceedings of the 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'15)*, pages 379–390, 2015.

[31] F5 Networks, Inc. 2016 DDoS Attack Trends. November 2016.

[32] Darren Anstee, Paul Bowen, C.F. Chui, and Gary Sockrider. Worldwide Infrastructure Security Report. Arbor Networks, Inc., 2016.

[33] Martin McKeay et al. The Q4 2016 State of the Internet / Security Report. Akamai, 2017.

[34] DDoS Threat Landscape Report 2015-2016. Imperva, Inc., August 2016.

[35] Arne Welzel, Christian Rossow, and Herbert Bos. Delving into Internet DDoS Attacks by Botnets: Characterization and Analysis. In *Proceedings of the 7th European Workshop on System Security (EuroSec'14)*, pages 3:1–3:6, 2014.

[36] Arman Noroozian, Maciej Korczyński, Carlos Hernandez Gañan, Daisuke Makita, Katsunari Yoshioka, and Michel van Eeten. Who gets the boot? analyzing victimization by ddos-as-a-service. In *Proc. of the 19th International Symposium on Research in Attacks, Intrusions, and Defenses (RAID 2016)*, 2016.

[37] Stephanie Weagle. Financial Impact of Mirai DDoS Attack on Dyn Revealed in New Data. https://www.corero.com/blog/797-financial-impact-of-mirai-ddos-attack-on-dyn-revealed-in-new-data.html, February 2017.

[38] Matthew Prince. Technical Details Behind a 400Gbps NTP Amplification DDoS Attack. https://blog.cloudflare.com/technical-details-behind-a-400gbps-ntp-amplification-ddos-attack/, February 2014.

[39] Swati Khandelwal. 602 Gbps! This May Have Been the Largest DDoS Attack in History. http://thehackernews.com/2016/01/biggest-ddos-attack.html, January 2016.

[40] Sharad Agarwaly, Travis Dawson, and Christos Tryfonasy. DDoS Mitigation via Regional Cleaning Centers. Sprint ATL Research Report RR04-ATL-013177, January 2004.