NBER WORKING PAPER SERIES

EVIDENCE OF DECREASING INTERNET ENTROPY: THE LACK OF REDUNDANCY IN DNS RESOLUTION BY MAJOR WEBSITES AND SERVICES

Samantha Bates
John Bowers
Shane Greenstein
Jordi Weinstock
Yunhan Xu
Jonathan Zittrain

Working Paper 24317
http://www.nber.org/papers/w24317

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2018, Revised March 2018

The authors would like to thank Hans Christian Gregersen and Matt Phillips for their work in preparing early versions of our dataset. The authors would also like to thank David Dinin and Andy Ellis for lending us their expertise in DNS and other technical matters. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w24317.ack

Evidence of Decreasing Internet Entropy: The Lack of Redundancy in DNS Resolution by Major Websites and Services

Samantha Bates, John Bowers, Shane Greenstein, Jordi Weinstock, Yunhan Xu, and Jonathan Zittrain

NBER Working Paper No. 24317
February 2018, Revised March 2018
JEL No. L2,L22,L86

ABSTRACT

This paper analyzes the extent to which the Internet's global domain name resolution (DNS) system has preserved its distributed resilience given the rise of cloud-based hosting and infrastructure. We explore trends in the concentration of the DNS space since at least 2011. In addition, we examine changes in domains' tendency to "diversify" their pool of nameservers – how frequently domains employ DNS management services from multiple providers rather than just one provider – a comparatively costless and therefore puzzlingly rare decision that could supply redundancy and resilience in the event of an attack or service outage affecting one provider.

Samantha Bates
Harvard Law School
Cambridge MA 02138
sbates@law.harvard.edu

Jordi Weinstock
Harvard Law School
Cambridge MA 02138
jweinstock@law.harvard.edu

John Bowers
Harvard Law School
Cambridge MA 02138
johnbowers@college.harvard.edu

Yunhan Xu
Harvard Law School
Cambridge MA 02138
yunhanx@gmail.com

Shane Greenstein
Technology Operation and Management
Morgan Hall 439
Harvard Business School
Soldiers Field
Boston, MA 02163
and NBER
sgreenstein@hbs.edu

Jonathan Zittrain
Berkman Klein Center for Internet
and Society, and
Harvard School of Engineering
and Applied Sciences Cambridge MA 02138
a2jz@law.harvard.edu

Evidence of Decreasing Internet Entropy: The Lack of Redundancy in DNS Resolution by Major Websites and Services.

Samantha Bates, John Bowers, Shane Greenstein, Jordi Weinstock, Yunhan Xu, and Jonathan Zittrain[1]

Abstract

This paper analyzes the extent to which the Internet's global domain name resolution (DNS) system has preserved its distributed resilience given the rise of cloud-based hosting and infrastructure. We explore trends in the concentration of the DNS space since at least 2011. In addition, we examine changes in domains' tendency to "diversify" their pool of nameservers -- how frequently domains employ DNS management services from multiple providers rather than just one provider -- a comparatively costless and therefore puzzlingly rare decision that could supply redundancy and resilience in the event of an attack or service outage affecting one provider.

Introduction

On October 21, 2016, it appeared to many worldwide Internet users that the network had broken. For example, many in North America or Europe that day trying to access major websites such as Netflix, CNBC, and Twitter, received only blank screens or error messages. The cause was a distributed denial of service (DDoS) attack on Dyn, a major domain name system service provider. Dyn reported that its servers were overwhelmed by a flood of requests from a botnet of "Internet of Things" (IoT) devices infected by Mirai malware.[2]

As a DNS provider that might be selected by a website, Dyn enables Internet traffic by translating the site's domain name (URL) into the IP address where the server behind that domain is to be found. During the attack, Dyn servers were unable to process users' translation requests (both legitimate and illegitimate), and as a result, users lost access to web domains contracting with Dyn such as Netflix, CNBC, and Twitter.[3] While it is impossible to calculate the economic cost of such an attack, Dyn itself claims that "[t]he cost of a DDoS

[1] The authors would like to thank Hans Christian Gregersen and Matt Phillips for their work in preparing early versions of our dataset. The authors would also like to thank David Dinin and Andy Ellis for lending us their expertise in DNS and other technical matters.

[2] Scott Hilton, "Dyn Analysis Summary of Friday October 21 Attack" (October 26, 2016) http://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack/. Archived at https://perma.cc/YW5C-MDEV

[3] Lily Hay Newman, "What We Know About Friday's Massive East Coast Internet Outage," *Wired* (October 21, 2016) https://www.wired.com/2016/10/Internet-outage-ddos-dns-dyn/. Archived at https://perma.cc/3BU2-6F4K

attack can quickly reach up to million dollars an hour in lost revenue and productivity for the average size company."[4]

The Dyn attack illustrates how fragile the Internet can be when malicious actors know how to exploit its vulnerabilities. The "entropy" of the Internet – the fact that it comprises many servers, and they are traditionally found in a variety of physical and logical locations – makes it more resilient to random errors, but it remains vulnerable to targeted attacks against infrastructural components that have become centralized either by design or by practice.[5] The attack raises the questions of how necessary any existing centralization may be; what the trade-offs are of seeking to decentralize it; and what baseline changes in entropy have occurred over the years. This study hazards answers to each of these questions.

We find an increasing concentration of DNS services in a small number of dominant cloud services companies. Coupled with domains' apparent tendency not to employ DNS services from multiple DNS providers, this concentration could pose a fundamental threat to the distributed resilience of the Internet. Our results also suggest ways to mitigate these issues.

What is DNS?

DNS servers perform a variety of functions that make them an integral part of the Internet's infrastructure. This paper will focus primarily on DNS's role as a "website directory" or authoritative resolver which translates a website's human-friendly domain name (ex. www.example.com) into a machine-friendly IP address (ex. 192.0.2.1) pointing to the location of the website's host on the Internet. The information needed to complete these translations is stored in definitive form by a domain's "authoritative nameservers" in files called "resource records" (RRs).[6]

DNS is structured as a hierarchically distributed database. At the very top of the hierarchy are 13 "root" servers.[7] Administered by a range of organizations including governments and branches of the US military, these root servers are responsible for storing RRs corresponding to Top-Level Domain (TLD) nameservers. A vast range of TLDs (including ".com," ".net," ".org," and country-level identifiers such as ".uk" or ".fr") can be found at the rightmost end of the URL addresses we use every day. Each TLD nameserver – or network thereof – is responsible for keeping RRs corresponding to the authoritative nameservers[8] of domains

---

[4] https://dyn.com/ddos/. Archived at https://perma.cc/94ED-SZ9D

[5] Réka Albert, Hawoong Jeong, and Albert-László Barabási, "Error and attack tolerance of complex networks" Nature 406, 378-382 (July 27, 2000) http://www.nature.com/nature/journal/v406/n6794/full/406378a0.html.

[6] There are many different types of resource records, but for the sake of simplicity we do not distinguish between them. You can see the full list of RR types here: https://en.wikipedia.org/wiki/List_of_DNS_record_types (Archived at https://perma.cc/3AQL-NFZ2).

[7] More accurately, 13 networks of root servers with internal redundancy.

[8] In practice, there can be additional non-authoritative nameservers between the TLD nameserver and the domain's authoritative nameserver. In such cases, the hierarchy is simply extended accordingly. For example, when requesting the IP address of https://www.oii.ox.ac.uk/, the resolver may be directed to the nameservers for 'uk', 'ac' and 'ox' before it reaches the domain's authoritative nameserver.

that fall within that TLD.[9] For example, the ".com" nameservers keep RRs for the authoritative nameservers for domain namespace encompassing domain names such as "www.google.com" and "www.amazon.com." Authoritative nameservers for a given domain namespace can be administered by that domain's owners, or management can be outsourced to one or more external providers such as Dyn, AWS, or Cloudflare. They form the last stage in the DNS hierarchy, storing RRs that provide translations between domain names and IP addresses.

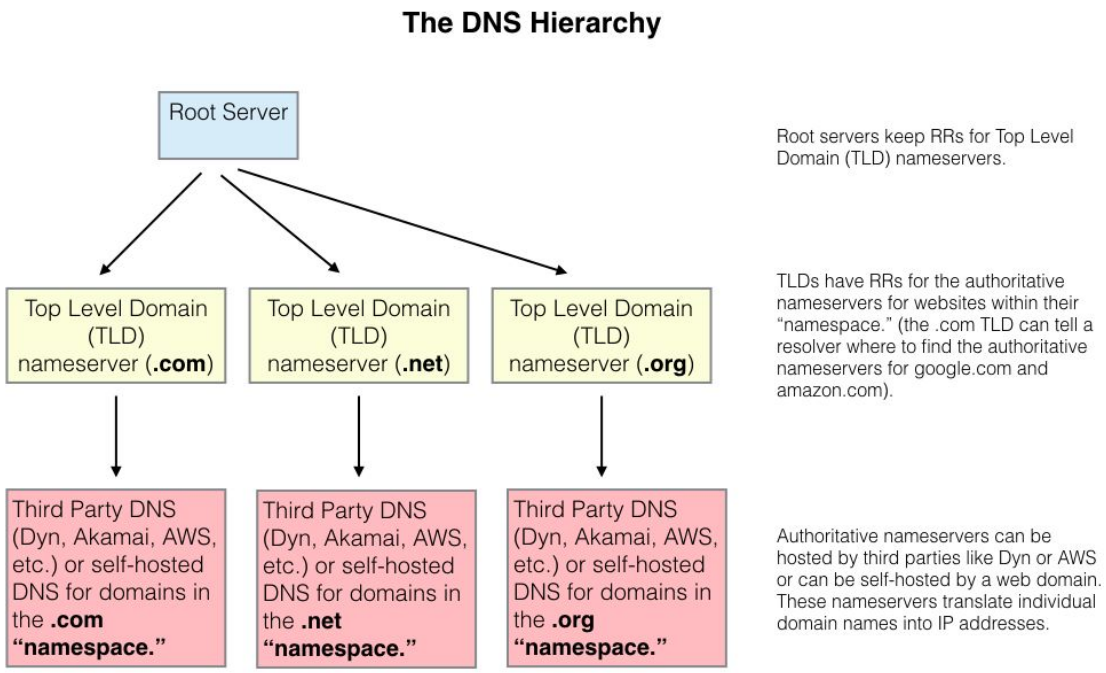

Figure 1

Tracing a DNS Request

To understand how the DDoS attack on Dyn interfered with this process, let's trace through the steps of a DNS lookup. When an application (such as a web browser) wants to access a page or resource located at a known domain name, it can leverage the DNS system to find a corresponding IP address. In principle, the application submits a request to a DNS "resolver" asking for the IP address corresponding to a given domain name, specified in URL format (ex. "www.google.com"). The resolver traces through this URL's period-separated components from right to left in order to zero in on the desired authoritative nameserver.

The resolver first queries a root nameserver, which replies with RRs corresponding to the TLD nameserver specified by the domain name (ex. ".com"). The resolver then queries that specified TLD nameserver with the second component of the domain name (ex. "google"). The TLD nameserver retrieves the RRs corresponding to that domain's authoritative nameservers (ex. "ns1.google.com") and returns them to the resolver. Finally, the resolver queries one of the authoritative nameservers and receives a usable IP address for

[9] These sets of RRs corresponding to a particular domain are often referred to as "NS records."

the domain. The IP address is passed back to the original application, which can use it to connect to the desired host. This entire process generally takes just milliseconds to finish.

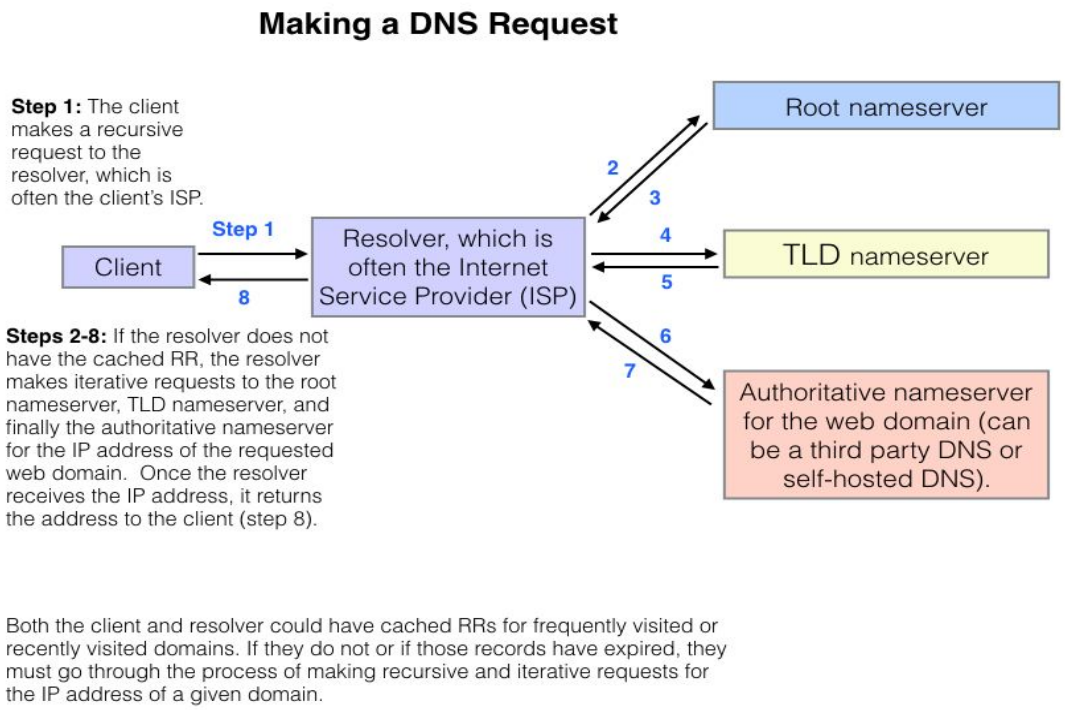

Figure 2

If every DNS request passed through all of these intermediary stages, the DNS infrastructure would face a constant and potentially overwhelming deluge of traffic. Luckily, RRs are cached at numerous points in the lookup process. Applications and Internet-connected devices maintain small caches of RRs for recently visited sites, as do Internet Service Providers (ISPs) and other intermediate nameservers. Of course, the content of these records can become inaccurate over time as IPs change. As such, RRs eventually expire. Most RRs must be replaced after twenty-four hours, though many larger sites implement much shorter expiration periods.[10] The duration for which nameservers are allowed to cache a RR without updating it is that RR's 'time to live' (TTL). An RR with a short TTL will have to be updated more often, but it is also more likely to remain accurate.[11]

If the authoritative nameserver corresponding to a particular domain name goes down (as many of those administered by Dyn did in the October DDoS attack), DNS resolvers become unable to update RRs that

[10] For example, the dig command, a tool for querying DNS servers (see https://en.wikipedia.org/wiki/Dig_(command) [Archived at https://perma.cc/JT2H-XE6M]), reveals that the RRs for www.spotify.com had a TTL of 150 seconds as of July 12th, 2017. View what a dig command looks like here: https://www.madboa.com/geek/dig/ (Archived at https://perma.cc/W56E-KGYT)

[11] Liu and Albitz, *DNS and BIND*, chapter 2, section 7, "Caching." See also Liu and Albitz, *DNS and BIND*, chapter 8, section 4, "Changing TTLs."

have expired or changed. In the event of a DNS failure IP addresses may still resolve to the correct host, but it is very unlikely that any user will know the exact current IP addresses of the websites they want to access. As such, a DNS malfunction effectively prevents users from accessing the content they have requested even if that content is hosted on an otherwise healthy server.

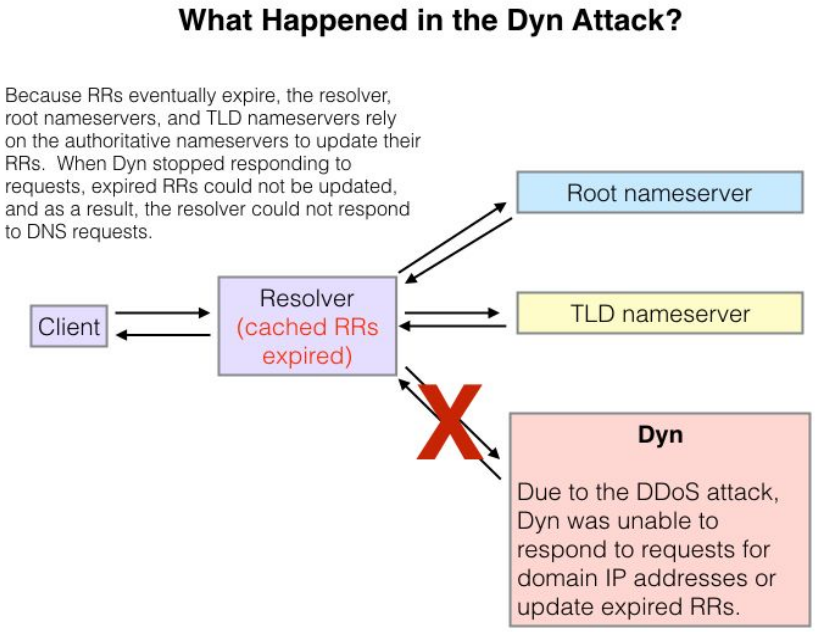

Figure 3

DNS: Designed for Resilience?

As is evident from the hierarchical architecture outlined above, the global Domain Name System (DNS) infrastructure is distributed by design.[12] Rather than being handled by a single master server mapping domain names to IP addresses, DNS lookups rely on interactions among millions of different servers worldwide. This distributed model theoretically brings a degree of segmentation and redundancy to the DNS system, minimizing "single points of failure" at which technical breakdowns result in access problems for significant swaths of the Internet.[13]

However, the rise of cloud-based hosting and domain management services threatens to overturn this distributed model of resilience. Companies such as Amazon Web Services[14], Akamai[15], and Dyn[16] offer scalable and often easily configurable external DNS hosting options alongside other cloud services, making it easier than ever to offload DNS management. External DNS hosting can offer significant advantages in terms of load balancing, reliability, and geographic reach. Dyn's DNS service boasts points of presence on

[12] https://cseweb.ucsd.edu/classes/wi01/cse222/papers/mockapetris-dns-sigcomm88.pdf (Archived at https://perma.cc/L46L-LG4W)
[13] http://www.bau.edu.jo/UserPortal/UserProfile/PostsAttach/10617_1870_1.pdf
[14] https://aws.amazon.com/route53/. Archived at https://perma.cc/K7YT-PVVM
[15] https://www.akamai.com/us/en/products/cloud-security/fast-dns.jsp. Archived at https://perma.cc/9YKJ-FQL8
[16] https://dyn.com/dns/. Archived at https://perma.cc/94ED-SZ9D

five continents, a team of dedicated security experts, an enormous performance analytics engine, and a multitude of load balancing and traffic steering features.[17]

Alongside these advantages, however, the consolidation of DNS services for a vast range of web domains into the hands of a relatively small number of providers could potentially pose a threat to the stability of the Internet. With consolidation comes single points of failure that create opportunities for simultaneous downtime. As in the case of the Dyn attack, the reachability of many domains can easily hinge on the resilience and stability of a single monolithic provider. Moreover, if the DNS market moves towards an oligopolistic or monopolistic structure, the lack of competition among DNS providers may decrease the diversity of providers available to consumers. Such a market shift could push consolidation even further while stifling consumer-friendly business practices and making DNS servers even more attractive targets to malicious actors.

The fact that externally hosted DNS providers offer high quality service while concentrating the DNS space presents an important tradeoff between two different patterns of downtime. On one hand, having a relatively small number of large externally hosted DNS providers with significant market share generates single points of failure which expose large segments of the Internet to simultaneous downtime. On the other, externally hosted DNS providers are – as mentioned above – generally very successful in maintaining almost perfect uptime while provisioning a range of sophisticated additional services that can substantially improve DNS performance. A tradeoff thus emerges between occasional periods of widespread downtime and potentially more frequent but highly distributed patterns of DNS downtime for individual sites or smaller externally-hosted DNS providers.

The Stakes of DNS Resilience

The Dyn attack provides a vivid illustration of how DNS infrastructure vulnerabilities – and DNS space concentration – can wreak havoc on the stability of the Internet. Prompting a widespread Internet outage that left many high-profile websites inaccessible for a period of hours,[18] the attack's devastating success highlights many of the ways in which a concentrated DNS space with relatively little provider diversification on the part of domain administrators can leave even large firms vulnerable to service disruptions.

Why, though, does DNS downtime matter? Is it really a big deal if a swath of the Internet becomes inaccessible for a few hours every once in awhile? In short, yes. Even brief periods of downtime can have a dramatic impact on the economic well-being of affected Internet companies. A 2014 survey[19] of 270 North American companies with a significant web presence found that about 50% of DDoS attacks cost their targets more than $20,000 an hour, with 49% of attacks lasting between six and twenty-four hours. Some

---

[17] https://dyn.com/dns/managed-dns/. Archived at https://perma.cc/BUU6-BSAK. In addition to translating domain names into IP addresses, DNS servers can also be used to direct and balance Internet traffic so that no individual server is burdened by too many requests. These types of functions are called load balancing and traffic steering. For the purposes of this paper we investigate the impact of concentration only on DNS's role as a "website directory." See Liu and Albitz, *DNS and BIND*, chapter 10, section 7, "Round-Robin Load Distribution."

[18] https://www.nytimes.com/2016/10/22/business/Internet-problems-attack.html

[19] https://lp.incapsula.com/rs/incapsulainc/images/eBook%20-%20DDoS%20Impact%20Survey.pdf. Archived at https://perma.cc/9RDY-FR8R

larger companies can suffer more than $100,000 an hour in IT, security, and sales related damages. About 87% of companies that had been targeted by a DDoS attack reported "at least one non-financial consequence such as loss of customer trust, loss of intellectual property, and virus/malware infection."[20]

DNS downtime has a range of other less explicitly economic impacts on a company's web presence. Search engines use web crawlers, automated scripts that can be used to copy visited websites or search for specific information, to index pages and determine search result presentation precedence.[21] If a site is unreachable during a crawl due to DNS downtime, the crawler will mark it as having been "uncrawlable" at the time of attempted access – a designation with potentially devastating consequences for a site's visibility.[22] Furthermore, since DNS is involved in handling services such as a domain's email functionality, DNS downtime can cripple communications within and between companies.

The DNS infrastructure will face more pressure from malicious actors in the coming years, not less. An annual report[23] published by the cybersecurity company Nexusguard found that in Q1 2017 the number of DNS attacks they observed "registered a 380% year-on-year growth, suggesting that DDoS attacks occurred more frequently than the same period a year ago."[24] The report partially attributes the proliferation of massive DDoS-capable botnets to the rise of the so-called "Internet of Things," which continues to bring an enormous number of often poorly secured web-connected devices online. The "Mirai" botnet which brought Dyn to its knees is only one example of such a weapon. As more and more devices are added to the global network – with billions projected to be added each year[25] – the number and power of botnets will almost certainly increase. DNS resilience is not simply about protecting against today's DDoS attacks – it is also about anticipating and preparing for tomorrow's.

Of course, it doesn't always take hundreds of thousands of unsecured webcams and Internet-connected toasters or other IoT devices to bring down a cloud services behemoth. In the course of routine maintenance in early 2017, a tech at Amazon accidentally caused a number of Amazon web servers to go offline. The server failure impacted services like Slack, Quora, and Medium that rely on Amazon cloud storage.[26]

---

[20] https://lp.incapsula.com/rs/incapsulainc/images/eBook%20-%20DDoS%20Impact%20Survey.pdf. Archived at https://perma.cc/9RDY-FR8R

[21] Web crawlers are automated scripts or programs that search the web for specific information (usually for indexing purposes) in a systematic way. See https://www.sciencedaily.com/terms/web_crawler.htm (Archived at https://perma.cc/UF2E-3W7B?type=image) and https://en.wikipedia.org/wiki/Web_crawler (Archived at https://perma.cc/B66K-2JDK).

[22] https://www.dosarrest.com/ddos-blog/how-ddos-attacks-can-impact-your-seo/. Archived at https://perma.cc/E6G5-8KNQ

[23] https://www.nexusguard.com/hubfs/Nexusguard_DDoS_Threat_Report_Q1_2017_EN.pdf. Archived at https://perma.cc/2KW5-KU75

[24] https://www.nexusguard.com/hubfs/Nexusguard_DDoS_Threat_Report_Q1_2017_EN.pdf. Archived at https://perma.cc/2KW5-KU75

[25] http://spectrum.ieee.org/tech-talk/telecom/Internet/popular-Internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated. Archived at https://perma.cc/JA6K-XH72

[26] Jon Fingas, "Amazon outage breaks large parts of the Internet" *Engadget* (February 28, 2017) https://www.engadget.com/2017/02/28/amazon-aws-outage/ (Archived at https://perma.cc/BZ79-8689) and Timothy J. Seppala, "Amazon admits that a typo took the Internet down this week" Engadget (March 2, 2017) https://www.engadget.com/2017/03/02/amazon-admits-that-a-typo-took-the-Internet-down-this-week/. Archived at https://perma.cc/K3SZ-B9JA.

Amusingly, the AWS status page failed to properly reflect the outage due to hard dependencies on the service that it was designed to monitor.[27] With consolidation comes the risk that simple mistakes on the part of engineers and administrators will cascade into widespread outages.

Methodology

To analyze trends in the concentration and diversification of the DNS space over time, we sampled the top 1000 U.S. domains in the ".com," ".net," or ".org" Top-Level Domains (TLDs) according to Alexa Top Sites [28] listings on a monthly basis between November 2011 and May 2017. We recognize that our results could change if we included domains from other regions, such as ".cn" and ".ru," in our dataset. However, we chose to focus on traditional domains, ".com," ".net," and ".org" because they are among the longest standing top level domains and comparatively represent a broad spectrum of the Internet. Due to our sampling method, the set of domains examined each month varied depending on the composition of the Alexa rankings.[29] This sampling methodology is also approximate because our dataset is missing about 3.5% of rankings per month. The motivation for this approach is drawn from the use of market indexes in economics – indexes such as the S&P500[30] are composed to reflect the performance and behavior of a broader market. It is important to note that DNS providers' market share in this sample does not reflect that of the broader Internet – our focus is on high-traffic domains which are far less likely to use low-end market oriented providers such as GoDaddy DNS.[31]

---

[27] https://twitter.com/awscloud/status/836656664635846656?lang=en. Archived at https://perma.cc/5RD5-D97B

[28] http://www.alexa.com/topsites. Archived at https://perma.cc/644F-6NGN

[29] Alexa Top Sites data has some deficiencies, but was the best available dataset for our study. One drawback of the dataset is that Alexa's ranking algorithm is not transparent. See https://support.alexa.com/hc/en-us/articles/200449744-How-are-Alexa-s-traffic-rankings-determined. Archived at https://perma.cc/CF9U-Z9F8. However, it should be noted that other ranking companies also prefer to keep their ranking algorithms secret and it is often not possible to obtain verified data from any of these companies. Moreover, the top ranked sites identified by Alexa, Quantcast, Comscore, and other ranking companies are comparable, confirming that our dataset of top ranked sites is accurate. While there may be minor differences among rankings depending on the company, these differences tend to occur at the extreme end of samples (ranks beyond 700 to 8,000). Finally, as these discrepancies are random, they are immaterial from a statistical standpoint.

[30] http://www.indexologyblog.com/2013/07/09/inside-the-sp-500-selecting-stocks/. Archived at https://perma.cc/HS8A-W273

[31] https://www.datanyze.com/market-share/dns/

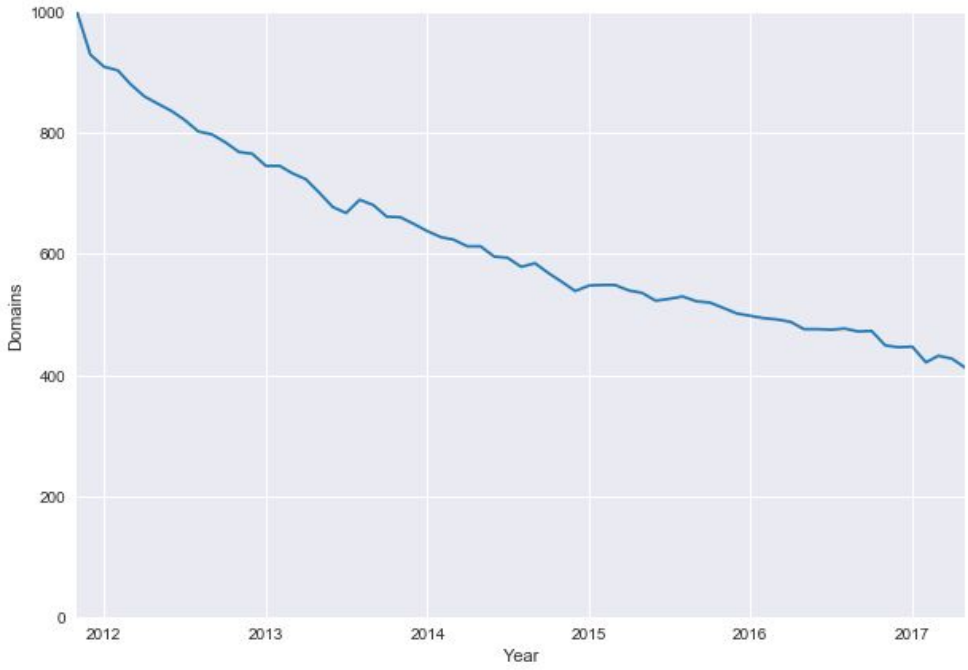

Figure 4: Original top 1000 domains in the top 1000 sites (ranked by Alexa) over time

It is also challenging to determine the importance and representativeness of a domain because it is not as easy to quantitatively model as that of a publicly traded entity. While the contributions of members of a conventional stock market index might be weighted using relatively straightforward and publicly available metrics such as market capitalization, weighting domains is not nearly so simple. Even seemingly basic metrics such as page views per month are very difficult to estimate accurately without insider analytics.[32, 33] However, despite these challenges, we believe that our data are representative of most high-traffic Internet sites within the TLDs examined. Given that "eyeball share" on the Internet is highly concentrated within a relatively small set of sites, sampling 1000 of the most active sites on the Internet every month captures a sample which represents a large proportion of all Internet behavior in the U.S..

We collected historical nameserver information for each of the domains identified in our dataset using an API offered by CompleteDNS. These data record changes in each domain's nameserver registration on a monthly basis, allowing us to identify migrations to and away from external DNS providers or self-hosted DNS solutions. We cross-checked our historical nameserver information against a number of comparable providers and found no evident discrepancies or errors. We were not concerned with how much of a domain's DNS traffic was routed through a given nameserver. Rather, we looked at the number of distinct authoritative nameservers registered by a domain and which providers managed those nameservers. By combining our monthly sample of domains with these historical nameserver data, we were able to generate a month-by-month timeline of nameserver configurations for our sample.

---

[32] https://moz.com/rand/traffic-prediction-accuracy-12-metrics-compete-alexa-similarweb/. Archived at https://perma.cc/32SC-MBEU.

[33] https://www.screamingfrog.co.uk/how-accurate-are-website-traffic-estimators/. Archived at https://perma.cc/24VC-WU52.

Concentration of the DNS Space

The extent to which a space of providers delivering vital services is concentrated has major implications for that space's vulnerability to catastrophic wide-ranging failure. If a large proportion of share in the space is divided among a small number of providers, outages affecting any one of those providers may have dramatic repercussions for the space as a whole. In other words, high-share providers have the potential to become single points of failure for large segments of the Internet. In examining the concentration of the DNS space, we aimed to shed light on the extent to which it falls prey to these vulnerabilities.

DNS as a Space

Our sampling methodology defines a set of high-traffic domains under the ".com," ".net," and ".org" TLDs. Within this set we can observe, using the methodology outlined above, continuities and changes in the DNS nameservers that domains identify, domains' DNS usage, and how domains adjust their DNS architectures, as well as the behaviors of new entrants into the sample. Each domain in our sample uses one or more DNS providers. Some domains host and manage their own DNS nameservers, in which case they are their own DNS provider. Others rely entirely on external providers such as Dyn, Amazon Web Services (AWS), and Akamai. At any given time, a domain may have multiple nameservers (generally 2-8) managed by a combination of providers or by a single provider.

We define the DNS space as the set of providers – including external DNS providers like Dyn, Akamai, and AWS, and others, as well as "self-hosting" domains – handling DNS for our sample of domains. A provider's share of the space is defined as the number of domains in the sample for which that provider handles at least one nameserver. The total size of the space is defined as the number of domains or "fractional domains"[34] administered by each provider summed across all providers present in the sample.

Concentration of the DNS Space as a Whole

To quantitatively model the concentration of the DNS space over time we adopted a metric from the antitrust economics literature: the Herfindahl-Hirschman Index (HHI). A standard measure of market concentration used by the Department of Justice, Federal Trade Commission, and Census bureau, the HHI is defined as the sum of squared market shares across all firms in a market.[35] It ranges from 0 (for a perfectly distributed market) to 10,000 (a perfectly concentrated market). More formally:

$$HHI = \sum_{i=1}^{N} s_i^2$$

Where N is the number of firms (DNS providers in our case) and s_i is the share for the *i*-th firm.

---

[34]In cases where a domain's DNS services is diversified among multiple providers, share is divided evenly among all providers used. For example, a domain managed by two providers would contribute 0.5 share to each. In economics, share in these cases is generally divided according to proportions of sales – a comparable division metric unfortunately does not exist in this case.

[35] http://heinonline.org/HOL/LandingPage?handle=hein.journals/fedred79&div=37&id=&page=. Archived at https://perma.cc/6G9J-5KKN.

The HHI is generally used to measure market share in terms of revenues collected by suppliers; however, that is not the case here. Because we are limited by the information that is publicly available (we do not have access to the pricing of services or the revenues collected by the suppliers in this market), the HHI in this paper measures market share based on supplier choice without accounting for revenue. Applying the HHI to our dataset revealed a considerable degree of consolidation within the DNS space between November 2011 and May 2017.

Figure 5: HHI of DNS providers in our dataset by year

As is apparent from figure 5 (pictured above), the HHI increased by a factor of about 6.9 over the observed timespan.[36] This indicates that a number of DNS service providers managed to significantly increase their proportional share of the DNS space in that timeframe, beginning to consolidate control of DNS services. The linearity of the trend is striking – gains in concentration have been relatively consistent in the long run despite yearly fluctuations.

To better understand the nature of this share capture, we looked at the percentage of share belonging to the top provider, top 4 providers, and top 8 providers in the DNS space over time (note that top providers occasionally changed on a month by month basis). As is reflected in figure 6 below, a small number of providers have come to largely dominate the DNS space. The percentage of share held by the top 8 providers more than doubled between November 2011 and May 2017, increasing from about 24% to about 59%. The percentage of share held by the top 4 providers grew by an even greater proportion, increasing from about 17% to about 49.8%. The top provider in the sample controlled 4.9% of share in November 2011 and 17.3% of share in May 2017.

[36] In recognition of the fact that a disproportionate volume of internet activity is dominated by a few major websites – Google, Facebook, Youtube, etc. – we recalculated the HHI using a weighting scheme so as to represent the outsized influence of these top websites. All domains' share in the sample were weighted by a factor of $w(R) = e^{2.6545 - \ln(R)}$ (where R is the domain's Alexa rank), a function derived from the regression of a domains' share of global traffic on its Alexa rank presented by Shiller et al. (DOI:10.3386/w23058). This weighting scheme had pronounced effects on the HHI: it grew from 795 in November 2011 to 964 in May 2017, experiencing frequent fluctuations due to small changes in the ordering of the top few sites. Dropping the top 5 domains from the dataset after weighting caused the HHI to grow from 262 in November 2011 to 566 in May 2017, and dropping the top 25 domains caused the HHI to grow from 138 in November 2011 to 619 in May 2017, much more in line with the unweighted trend illustrated above.

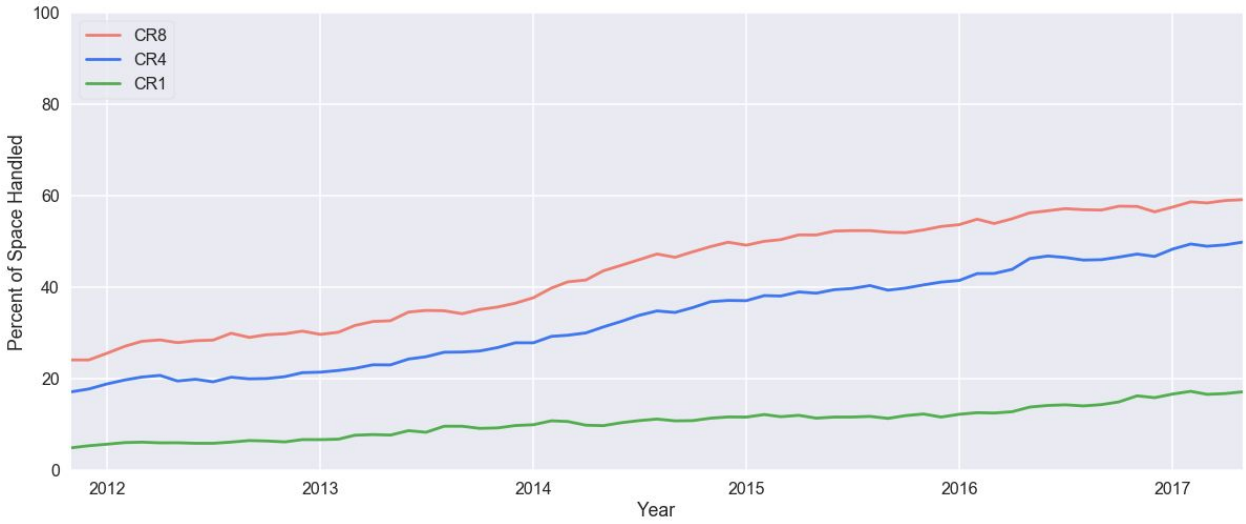

Figure 6: Percentage of the DNS market controlled by the top DNS providers

An analysis of share over time for several of these top providers offers further insight into changes in the DNS space. The figure below shows change in market share over time for all providers that were in the top 3 (ranked by market share) at any point in the sample timeline. Note that a number of rapidly expanding providers including Dyn, Akamai, AWS, and Cloudflare have captured a large proportion of the DNS space while other players such as Neustar and DNSPod have retained significant share over time. The massive expansion of AWS and Cloudflare (which collectively handle about a third of the entire space) is particularly striking, signalling the increasing influence of multi-service cloud-based platforms in the DNS space.

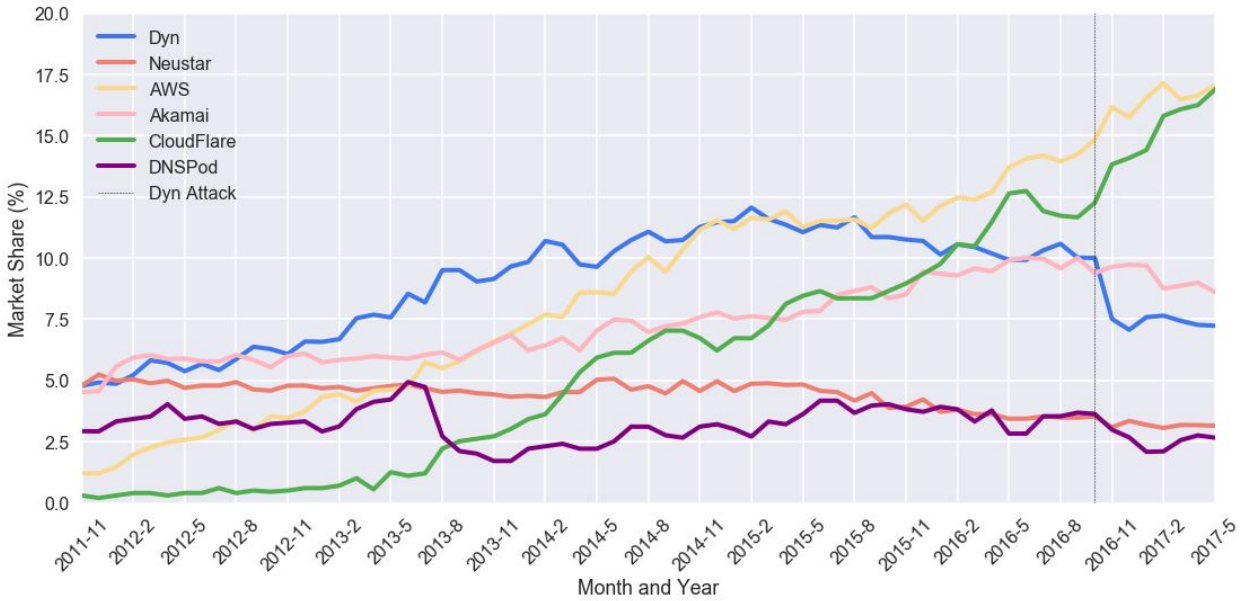

Figure 7: Market share of the top DNS providers

Breaking the space down into "original" domains (those that were present in the November 2011 sample) and "entrant" domains (those that were not) reveals a marked difference in the DNS hosting preferences of long

established domains and more recent ones. Entrant domains tended to use CloudFlare and AWS at much higher rates than original domains and used Akamai, Dyn, and Neustar relatively less than original domains. Cost may be one determinant of this divide – Akamai, for example, offers very expensive high performance DNS and hosting services that newer domain owners might be unable to afford. It is also worth noting that domains often use these providers for more than just DNS, as many also offer services including Content Delivery Network (CDN) assistance and site hosting.

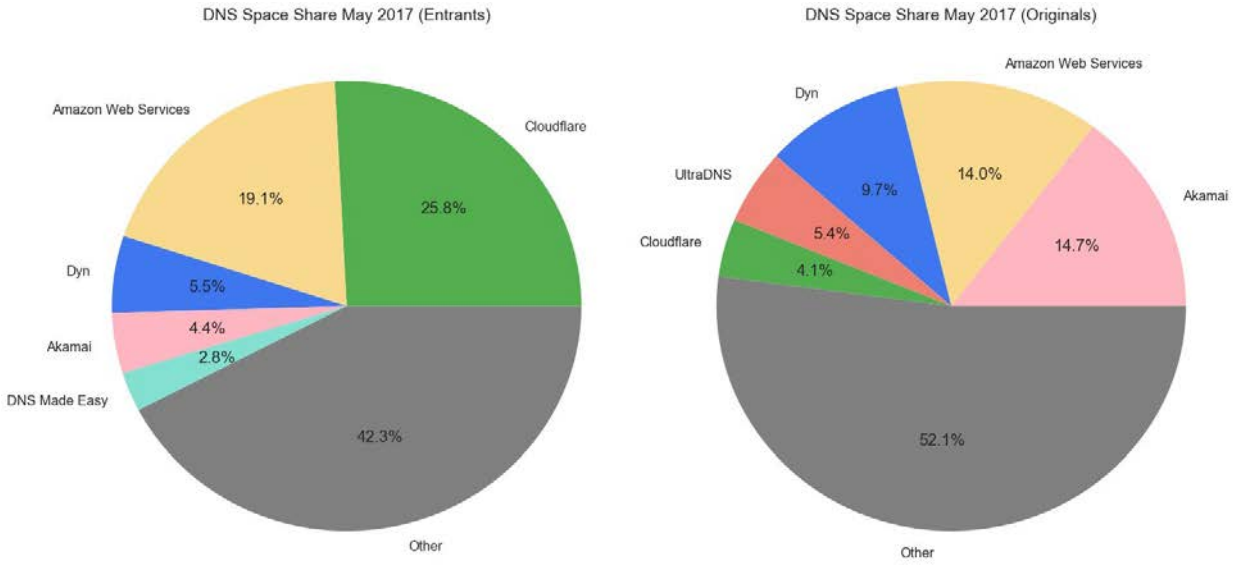

Figure 8: Original and entrant domain DNS provider preferences

The entrant domain space was also more concentrated than the original domain space, with a May 2017 HHI value of over 1100 as opposed to about 600 for the original domain space. This signals that newer domains which have broken into the high-traffic sphere may be relying more and more on a specific set of external DNS hosts.

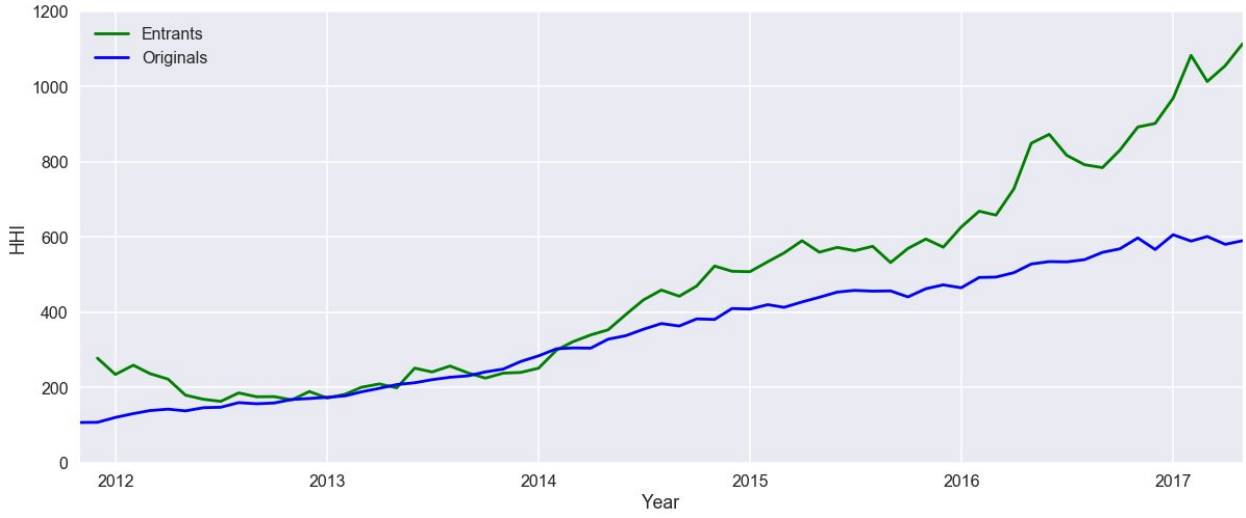

Figure 9: The HHI of original and entrant domains over time

Concentration of Self-Hosted vs. Externally Hosted DNS

While the HHI gives us a means of conceptualizing the DNS space in terms of existing market analysis frameworks, DNS is distinct from most conventional markets in that many domains choose to host and administer their own DNS nameservers rather than use external hosting providers such as Dyn or Cloudflare. Companies that host their own DNS nameservers may maintain a server on their property or host their nameservers with a service like Rackspace under their own names and remain responsible for managing it. By differentiating external hosting providers from self-hosting providers – domains administering their own DNS nameservers – we can both identify trends in the frequency of self-hosting and analyze the external hosting market independently.

Our analysis of self-hosting and external hosting reveals that concentration of the DNS space is driven by two primary forces – the concentration of the external hosting space into the hands of a small number of providers and a widespread migration away from self-hosted DNS towards externally hosted DNS. The rise of external hosting promotes concentration by enabling the consolidation of market share that would otherwise be distributed among a large number of self-hosts into the hands of a single entity.

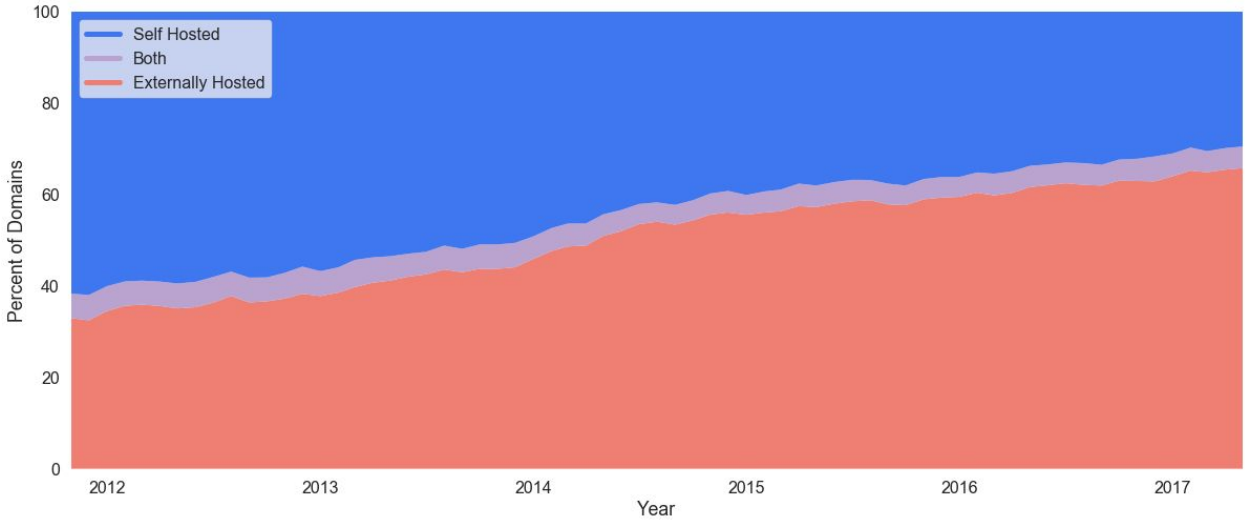

Figure 10: The percentage of domains that host their own DNS and/or rely on external DNS providers over time.

As is evident from figure 10 (above), external DNS hosting rapidly overtook self-hosted DNS in the period between November 2011 and May 2017.[37] The percentage of domains managed entirely by external DNS hosting providers grew from 32.9% to 65.7% over that period. By looking at the concentration of the two space segments independently, we can better pinpoint the sources of increased space-wide concentration. The HHI of the external hosting space more than doubled between November 2011 and May 2017 as a small

[37] Rerunning this analysis under the weighting scheme described in footnote 35 gave significantly different results: after weighting, the percentage of domains using externally hosted DNS services grew from 19.8% in November 2011 to 30.2% in May 2017. The lesser magnitude of this change is explained by leading – and heavily weighted – domains' proclivity towards self-hosting: Google's self-hosted domains account for about 25% of total weighted share in the sample throughout the timeline, but less than 1% of the unweighted sample.

number of providers took an increasing proportion of share. It also increased in the self-hosted DNS space. The HHI increased for the self-hosted segment of the DNS space because a number of large multiple domain self-hosting firms such as AOL and Alibaba, which have the ability to maintain their own DNS infrastructure, retained share while smaller firms switched to external hosting, shrinking the overall size of the self-hosting space. The external hosting space is massively more concentrated than the self-hosting space, with HHI averaging about 1116 over the timespan as opposed to about 43 for the self-hosting space.

Figure 11: The HHI of the external DNS hosting space.

The increased concentration of the DNS space can be said to be driven primarily by a combination of the consolidation of share in the external hosting space and the migration of domains out of the self-hosting space into the external hosting space. This shift is concurrent with a broader movement towards cloud-based site hosting and management platforms.

The Dyn Attack: Dangers of DNS Concentration

On October 21, 2016, the massive "Mirai" botnet (largely comprised of poorly secured Internet-connected devices) launched a sustained DDoS attack against systems critical to Dyn's DNS management services.[38] The attack took most Dyn services offline for hours, leaving domains reliant on Dyn for DNS services unreachable for many Internet users.

At the time of the attack, Dyn held about 10% share in the DNS space. Having expanded from holding 4.8% share in November 2011, Dyn was (and still is) among the biggest DNS providers in the sample and it greatly benefited from the increase in concentration of the market. Its clients included Twitter, Amazon, Zillow, SoundCloud, and eBay among many others.

[38] https://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack/. Archived at https://perma.cc/GE2P-CMKD.

Figure 12: Dyn's share of the DNS space over time

The dramatic consequences of the Dyn attack illustrate the extent to which DNS space consolidation has exposed large swaths of the Internet to shared single points of failure. When Dyn's systems were taken offline by the DDoS attack, they brought a sizeable fraction of the Internet down with them. Had fewer domains relied solely on Dyn's DNS services, the DDoS attack would have been far less catastrophic. If large external DNS hosting providers continue to absorb market share from smaller competitors and from the self-hosted segment of the DNS space, this non-distributed exposure will only become more pronounced.

Diversification of the DNS Space

While the concentration of the DNS space as outlined above has the potential to expose dangerous single points of failure in the DNS system, a powerful safeguard against such vulnerabilities is built into DNS itself. By allowing domains to register multiple nameservers, DNS gives potentially vulnerable domains the opportunity to diversify among multiple providers. For example, if you create a domain, www.example.com, you can choose to register more than one authoritative nameserver for that domain and each nameserver you register can be managed by a different provider. One of your nameservers could be managed by Dyn, example.Dyn.1, and a second could be managed by Akamai, example.Akamai.1. In the event of another DDoS attack on Dyn, users would still be able to access www.example.com as long as your DNS nameserver managed by Akamai remained up and running.

The importance of DNS redundancy and diversification is not a new realization. RFC 2182, last updated in 1997, outlines best practices for "selection and operation of secondary DNS servers." It explains that, "a major reason for having multiple servers for each zone is to allow information from the zone to be available widely and reliably to clients throughout the Internet, that is, throughout the world, even when one server is unavailable or unreachable."[39] Similarly, our findings suggest that by registering DNS servers managed by multiple DNS providers, a domain will remain accessible in the event that one provider fails due to an attack or technical malfunction because it is unlikely that both providers will go down at the same time. By choosing

[39] https://tools.ietf.org/html/rfc2182. Archived at https://perma.cc/7RS9-A2YK.

to work with multiple providers, any domain can therefore secure significant redundancy and robustness even in an increasingly concentrated DNS space.

Diversification of the DNS Space as a Whole

Our analysis showed that the majority of domains are not taking advantage of this opportunity for resilience through diversification. From November 2011 to October 2016, the proportion of domains in our sample using nameservers from just one provider fluctuated in the range between 91% and 93%. At no point in our timeframe did 1% or more of domains use 3 or more providers.

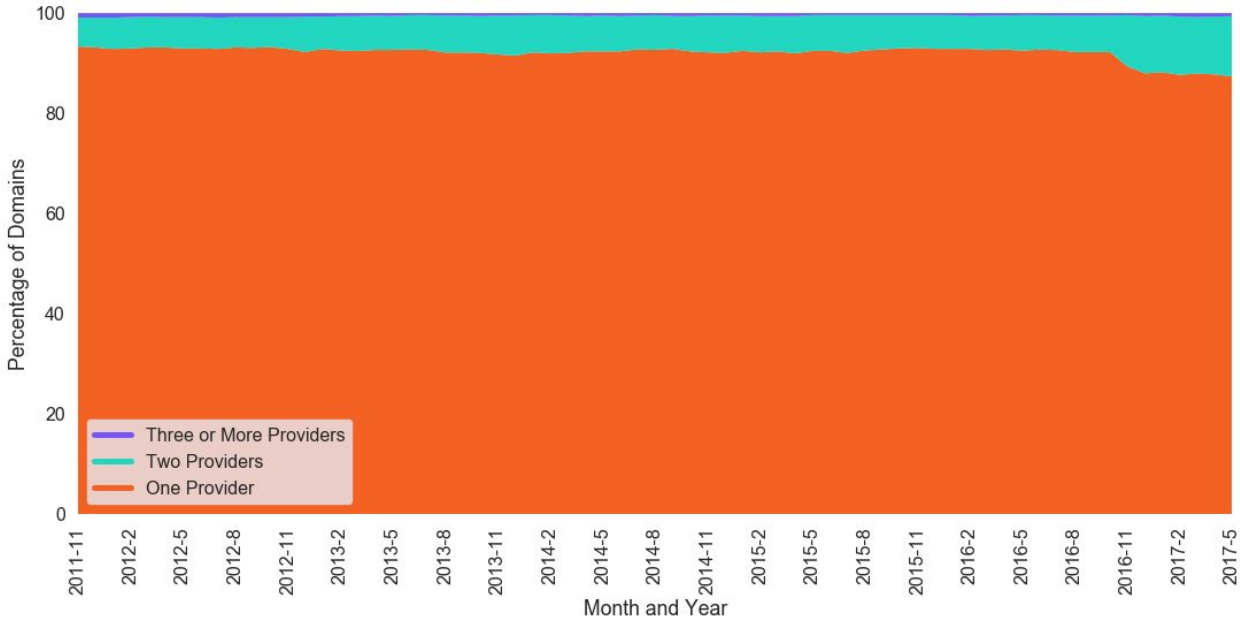

Figure 13: The percentage of domains that registered one, two, or three or more DNS providers

As the figure above illustrates, the Dyn attacks of October 2016 have seemingly spurred and sustained interest in DNS diversification. From October 1st, 2016 to November 1st, 2016, the percentage of domains using a single DNS provider fell from 92.2% to 89.4%. The percentage continued to decline between November 2016 and May 2017, reaching 87.3% in May 2017.

Interestingly, diversification was more common amongst the top ranked 100 domains per month in the sample. The percentage of domains using a single provider decreased from 95% in November of 2011 to 86% in October 2016. After the Dyn attack, it plummeted to a low of 78% in February 2017.

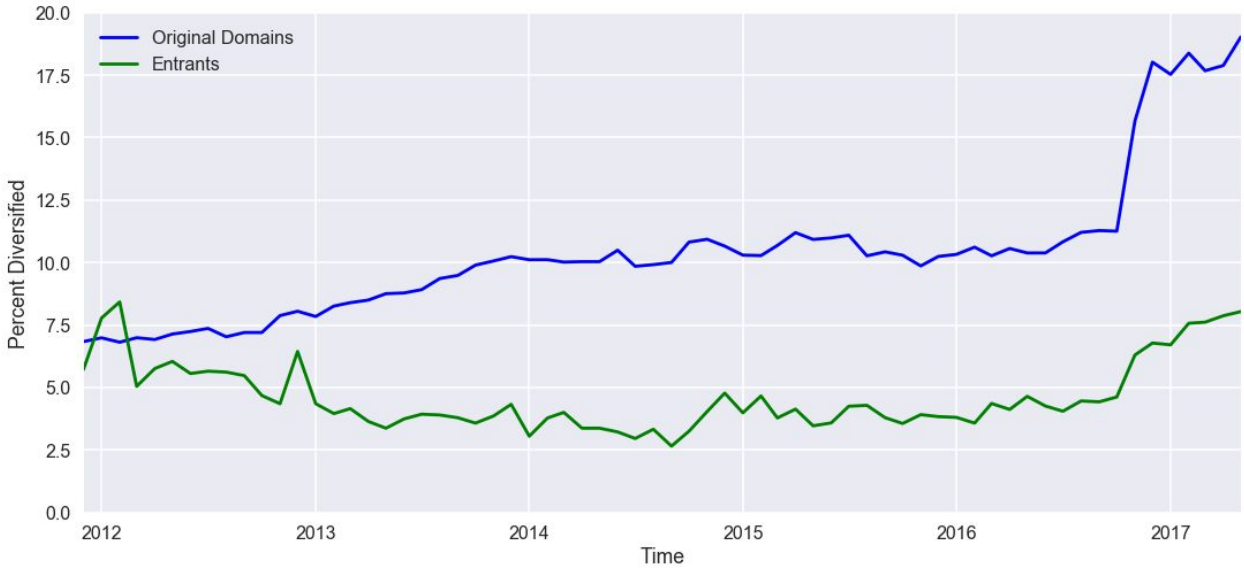

Figure 14: DNS diversification rates among original and entrant domains

This tendency to stick with one DNS provider was even more pronounced amongst entrant domains. As per figure 14 (above), entrant domains in the sample tended to diversify at considerably lower rates than original domains, with around 5% diversifying for most of the timescale prior to the Dyn attack. They also reacted less intensely to the Dyn attacks than did the original domains: their diversification rate increased by about half as much. This may signal that entrants are more wedded to single-provider cloud hosting. Such an explanation is particularly likely in the case of firms that developed since the explosive growth of the cloud began towards the beginning of our sample timescale. In contrast to the original domains that were created before there were many cloud based DNS providers and had to build their own DNS infrastructure, entrant domain owners have more external DNS providers to choose from. Moreover, external providers handle more than just DNS management, so choosing to host everything with one provider may be the easiest solution for entrants. For example, Cloudflare offers security features in addition to DNS management that make it impossible for domains to register DNS nameservers managed by other providers. Finally, entrant domain owners may be less risk averse because they have more to lose and less money at their disposal than larger, established domain owners.

Given the advantages of diversification, why would so many domains (entrants and established domain owners) elect not to use multiple DNS providers even in the aftermath of the revelatory Dyn attacks? Configuring DNS across multiple DNS providers can be a non-trivial technical undertaking. When a DNS server responds to a request by a DNS resolver, it authenticates its response with a set of "NS records," which must accurately reflect all of a domain's nameservers (even across multiple providers).[40] Using multiple DNS providers therefore requires that the domain's administrators be able to edit or otherwise synchronize NS records among DNS servers that are managed by different providers.

[40] https://tools.ietf.org/html/rfc2182. Archived at https://perma.cc/7RS9-A2YK.

Many large external DNS companies provide support and documentation for implementing DNS diversification, with some – such as Dyn – offering explicitly "secondary" DNS configurations as a service.[41] However, configuring multiple DNS providers can still require significant development efforts and the generation of (potentially insecure or buggy) bespoke code. *The Guardian* recently published a blog post documenting its efforts to register an additional DNS nameserver managed by AWS. The post describes the process of writing custom code to synchronize RRs between the primary and secondary DNS servers. For a company that does not have the same knowledge or technical resources, diversifying DNS providers may prove to be more trouble than it's worth.[42]

Moreover, it should be noted that large externally managed DNS providers – including Dyn – generally have near-perfect uptime records, with most attaining uptime of 99.9% or higher[43, 44] on a monthly basis. For many domains, particularly those with fewer network engineers at their disposal, externally managed DNS likely produces better results than self-hosted DNS with less potential for catastrophically long periods of downtime. The cost, difficulty, and potential technical issues associated with diversification among multiple providers may simply not be worthwhile for a majority of firms. As the saying goes, "Nobody ever got fired for buying IBM."[45] Additionally, as DNS providers continue to learn from experiences like the Dyn attack, these larger external DNS companies may accumulate knowledge that will help them prevent future failures. Some concentration of the market could in fact improve overall market practices, which would benefit all domain owners.

However, as the number of insecure web-connected devices continues to grow at a breakneck pace, it is likely that the scale and frequency of botnet-driven DDoS attacks capable of compromising even the largest providers will continue to increase.[46] If sophisticated externally managed DNS providers cannot scale their countermeasures accordingly, uptime figures for even the largest services may begin to decline. Furthermore, simultaneous downtime for large swaths of the web may be more destructive than more segmented periodical failures.

Types of Diversification: External and Self-Hosted DNS

When domains diversify, they may do so in a number of ways. For instance, a domain may choose to delegate DNS services between two or more externally hosted DNS providers. A firm working with Dyn might, for example, choose to add additional nameservers from Akamai or AWS to diversify their DNS. Alternately, a firm that manages its own DNS might choose to diversify with secondary services from a major externally hosted DNS provider or vice versa.

---

[41] https://help.dyn.com/using-external-nameservers/. Archived at https://perma.cc/M2LB-8E3V.
[42] https://www.theguardian.com/info/developer-blog/2016/dec/23/multiple-dns-synchronising-dyn-to-aws-route-53. Archived at https://perma.cc/WDP4-MY93.
[43] 99.9% uptime – "three nines" – is a common responsiveness goal for web domains
[44] http://www.thewhir.com/web-hosting-news/many-dns-cdn-services-attain-near-perfect-uptime-cloudharmony-report Archived at https://perma.cc/8KMZ-JECH.
[45] https://www.forbes.com/sites/homaycotte/2014/12/09/your-startup-delimma-nobody-ever-got-fired-for-buying-ibm/#3bdcb7fa16b6. Archived at https://perma.cc/PN7V-B5XF.
[46] http://www.americansecurityproject.org/the-rise-of-iot-botnets/. Archived at https://perma.cc/EMH7-CHL9.

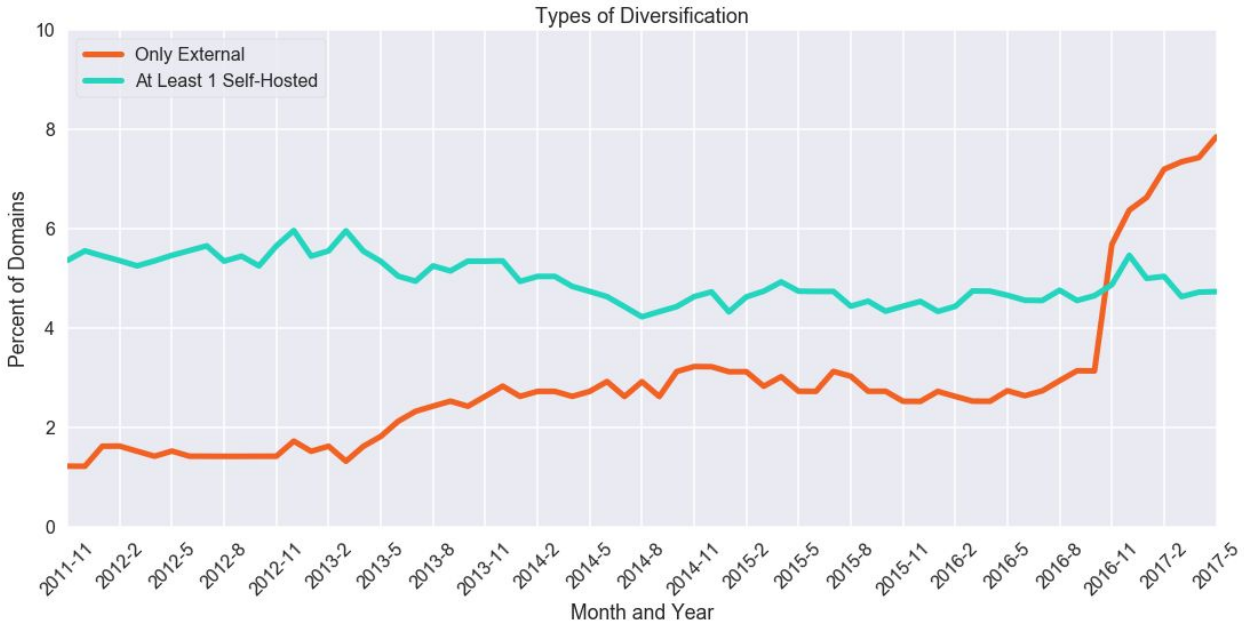

Figure 15: The proportion of domains that chose to self-host or employed only external DNS services over time

As is evident from the figure 15 (above), employing a combination of externally hosted DNS services and self-hosted DNS services was the most common form of diversification before the Dyn attack, after which diversification among external providers more than tripled. As of May 2017, about 62.4% of diversified domains opted to use multiple externally hosted DNS providers, while about 37.6% of diversified domains chose to employ a self-hosted domain nameserver and an externally hosted one.

There is a very significant degree of variation in patterns of diversification and growth among external DNS providers. As is evident from figure 16 (below) showing the 4 largest externally hosted DNS providers in the sample as of May 2017, customers of some externally hosted DNS providers tended to diversify much more than others. A fluctuating percentage of 81% and 89% of Akamai users remained undiversified throughout the sampled timescale as Akamai's DNS service has roughly doubled in size. Dyn's customer base (in terms of the number of domains supported), which almost tripled between November 2011 and early 2015, subsequently shrank to just over twice its size by May 2017. There was a steady increase in diversification rates among Dyn customers from November 2011 until the Dyn attack, at which point the percentage of undiversified domains using Dyn's services plummeted to 48%. The percentage of undiversified domains using Dyn's services decreased again to 36% in May 2017.

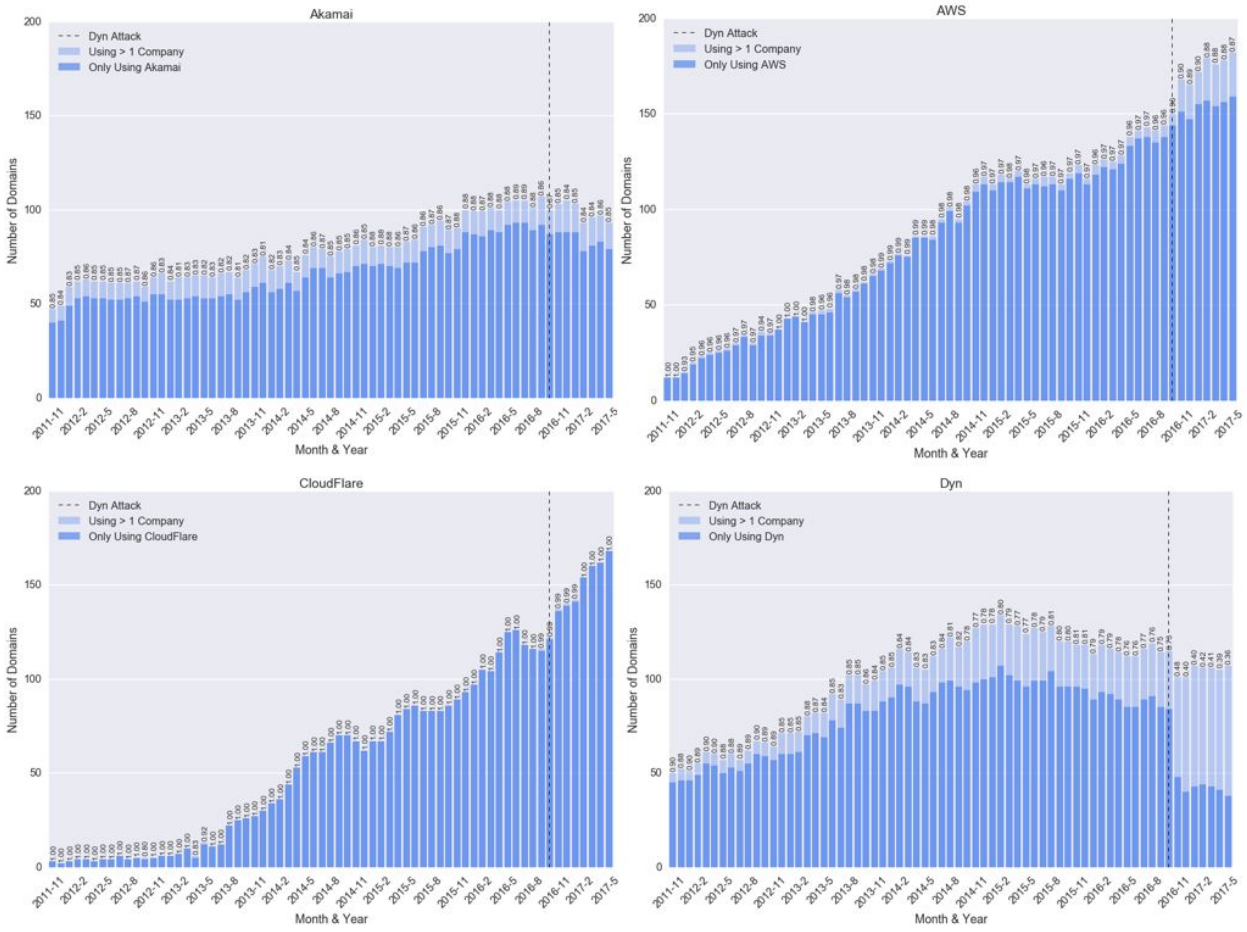

Figure 16: The number of undiversified and diversified domains using DNS services provided by one of the four largest external DNS providers: Akamai, AWS, Cloudflare, and Dyn.

Cloudflare maintained consistently near-zero diversification rates for the entire timescale, over the course of which it has exploded in size from having virtually no market share to becoming the second most dominant provider in the sample. This near-complete lack of diversification is a product of Cloudflare's security model, which requires that DNS traffic is routed through the Cloudflare network to protect against DDoS attacks and other network insecurities.[47] This approach does not allow domains to register a secondary nameserver managed by a different DNS provider. Until the Dyn attack (after which many domains added AWS as a secondary DNS provider) AWS also registered an extremely low diversification rate – no more than 4% of domains using AWS diversified at any point before the Dyn attack. Between the attack and May 2017, the proportion of diversified domains using AWS increased to about 13%.

[47] https://support.cloudflare.com/hc/en-us/articles/205195708-Step-3-Change-your-domain-name-servers-to-Cloudflare. Archived https://perma.cc/3FF7-P9BJ.

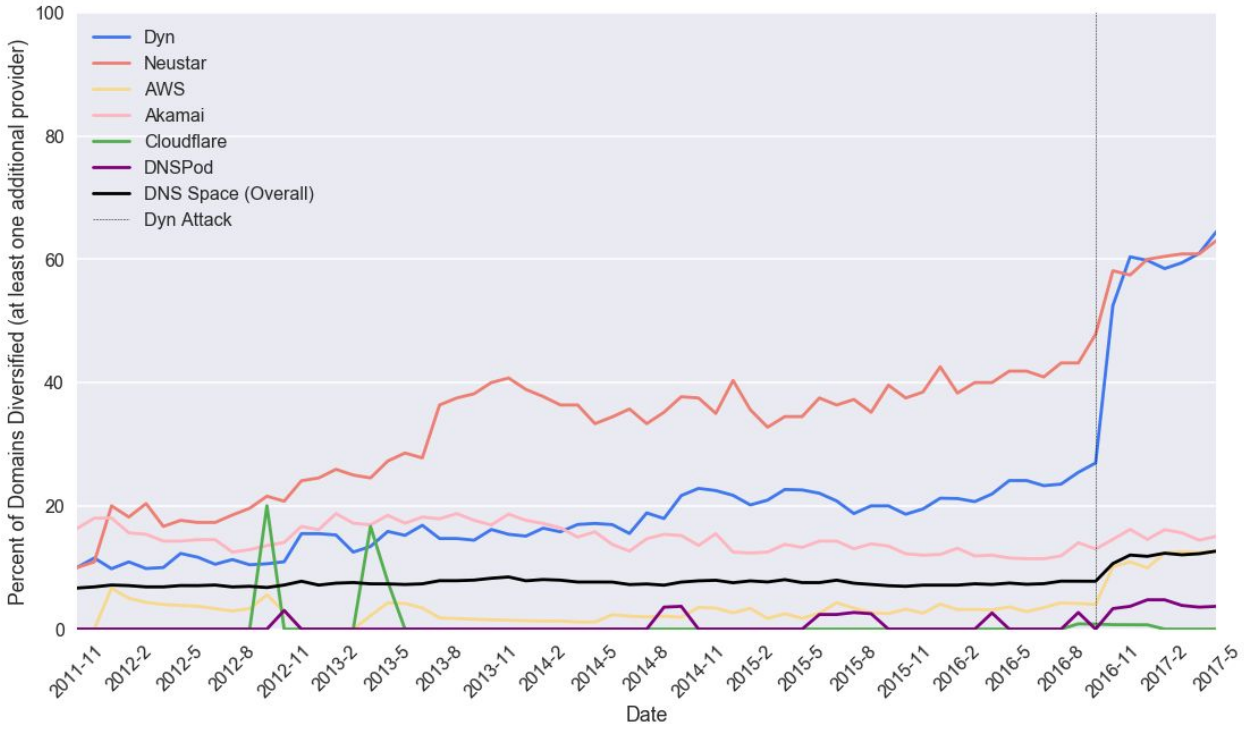

Figure 17: The percent of diversified domains before and after the Dyn attack.

These discrepancies likely point to differences in user experiences between external DNS providers which affect the ease with which domains can diversify among multiple DNS providers. It is worth noting that Dyn – which had the most diversified client domains of any large external DNS host following the October 2016 attack – provides specific guidelines for making its DNS services work smoothly with those of other providers.[48] Neustar[49] and DNS Made Easy,[50] the runners up to Dyn in this regard, also provide such instructions. AWS, which falls behind every other provider except for CloudFlare in terms of the diversification of its clients, offers no such readily accessible support.

The Dyn Attack: Diversification as a Corrective to Concentration

The Dyn attack highlighted the value of diversification among multiple DNS providers. At the time of the attack, 84 domains in our sample were using Dyn as their only DNS provider. Even more domains – including Netflix – were using Dyn as the sole DNS for alternate domains delivering content to their sites. The relationship between DNS vulnerability and CDNs should be explored further, but currently falls beyond the scope of this paper.[51]

[48] https://help.dyn.com/using-external-nameservers/. Archived at https://perma.cc/M2LB-8E3V.
[49] https://www.neustar.biz/blog/secondary-dns-service-paper. Archived at https://perma.cc/Q7VN-ZY24.
[50] https://www.dnsmadeeasy.com/services/secondarydns/. Archived at https://perma.cc/QQD4-P2UT.
[51] https://www.fastly.com/security-advisories/widespread-dyn-dns-outage-affecting-fastly-customers. Archived at https://perma.cc/P8B3-FD9N.

Following the Dyn attack, a number of major blogs and commentators offered vocal support for multi-provider DNS diversification.[52, 53, 54] A significant number of domains did diversify, bringing the percentage of domains in the sample using only one provider from 92.2% in October 2016 to 87.3% in May 2017. However, despite this coverage and evident shift in the DNS space, it seems that the lessons of the Dyn attack were learned primarily by those who suffered from them directly. Of the 52 domains that remained in the sample between October 2016 and May 2017 and diversified during that timeframe, the majority were using Dyn alone prior to the attack.

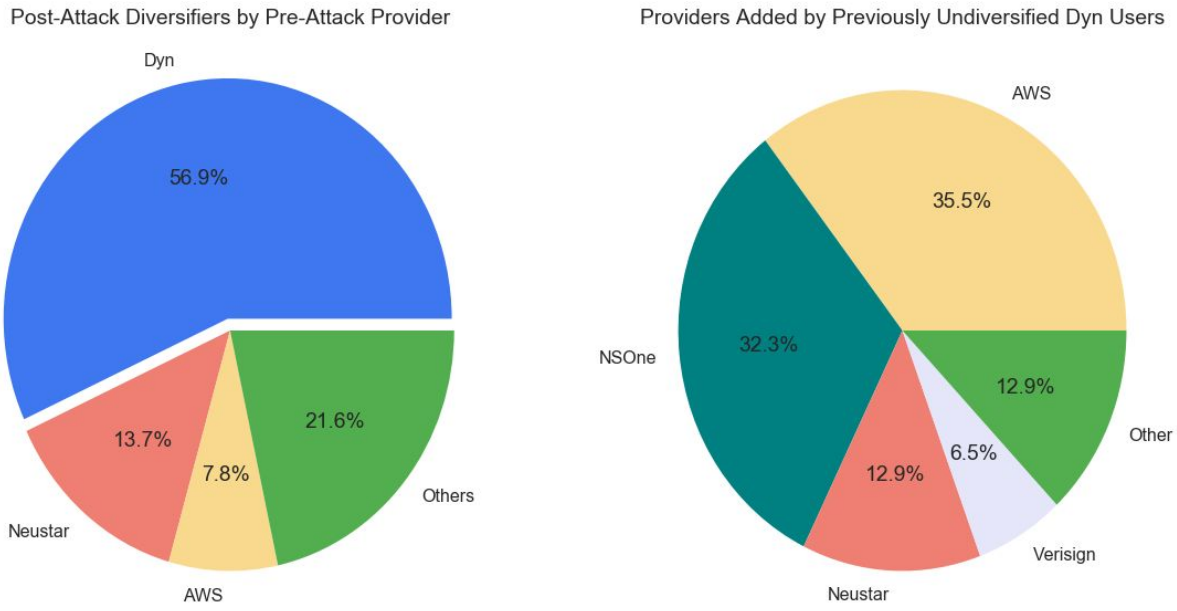

Figure 18: The percent of previously undiversified domains that registered additional DNS providers after the Dyn attack (graph on the left). The graph on the right shows the percent of Dyn users that employed additional DNS services from another externally hosted provider.

Of these Dyn users, more than two-thirds went to either AWS or NSOne, a relatively recent entrant into the DNS space which expanded its share somewhat in the wake of the Dyn attack. In accordance with the diversification patterns presented in figure 18, diversification through multiple externally hosted providers proved much more common than diversification through a combination of externally hosted and self-hosted DNS.

---

[52] https://www.internetsociety.org/blog/2016/10/how-to-survive-a-dns-ddos-attack-consider-using-multiple-dns-providers/. Archived at https://perma.cc/S3DM-TYXK.

[53] https://blog.thousandeyes.com/dyn-dns-ddos-attack/. Archived at https://perma.cc/P3UL-KH8F.

[54] https://security.radware.com/ddos-threats-attacks/threat-advisories-attack-reports/dns-services-under-attack/. Archived at https://perma.cc/W8LU-64MT.

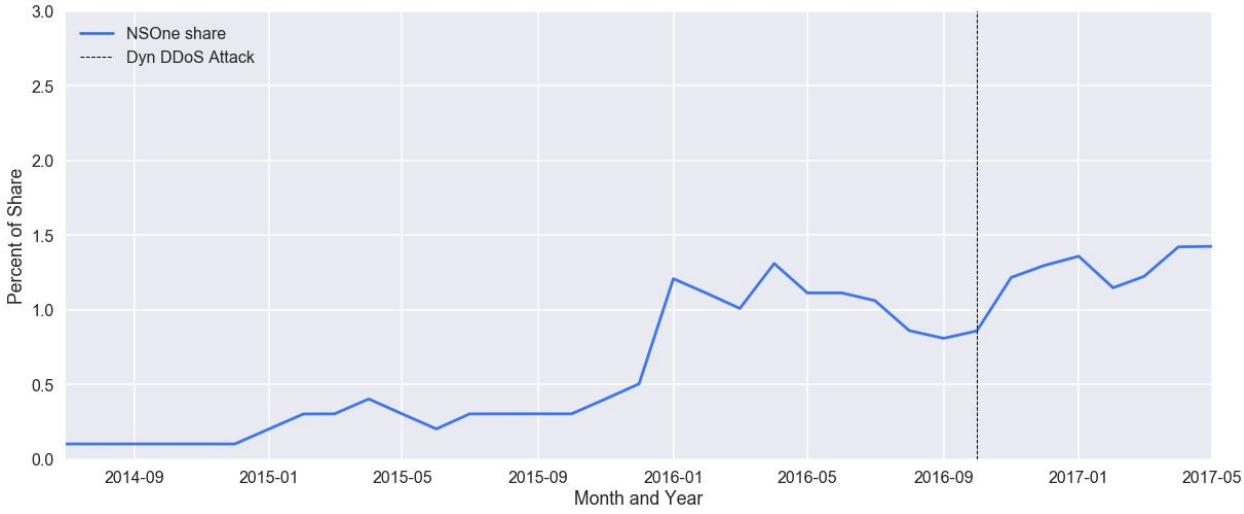

Figure 19: NSOne's market share of the DNS space before and after the Dyn attack

Conclusion

A stable DNS infrastructure is critical to the operation of the Internet. Growing cybersecurity challenges from botnets, ransomware, and other service disruption systems will stress-test even the biggest providers, and future downtime from such attacks is all but inevitable. With downtime comes SEO penalties and the loss of revenue and user trust. In addition to this threat of bad actors, even the largest cloud services providers have shown themselves to be vulnerable to large service outages.[55]

The concentration of the DNS space makes confronting these challenges all the more urgent. As a larger and larger proportion of the Internet's biggest sites fall under the management of a small number of externally hosted DNS providers, single points of failure will continue to emerge and grow in magnitude. While every company with an Internet presence will ultimately have to weigh its costs and benefits, diversification promises a powerful means by which many of the dangers of this new DNS environment can be mitigated. All would be prudent to consider altering their DNS architecture accordingly.

Ultimately, the biggest determinant of domain owners' willingness to diversify may be the extent to which providers choose to support and encourage such diversification. If single service lock-in is the model of cloud DNS platforms, widespread diversification is unlikely. However, providers could encourage diversification of DNS management services by requiring domain owners to select a secondary DNS provider or even specify which secondary providers they should choose. If more domains adopt and support affordable and easily configurable "Secondary DNS" models such as that offered by Dyn[56], a diversified and resilient DNS space could easily come into reach.

[55] http://www.americansecurityproject.org/the-rise-of-iot-botnets/. Archived at https://perma.cc/ATD3-9SLV.
[56] https://www.dyn.com/dns/secondary-dns/. Archived at https://perma.cc/7EF2-UX2H.