Victimization in DDoS Attacks: The Role of Popularity and Industry Sector

Muhammad Yasir Muzayan Haq[a,*], Antonia Affinito[a], Alessio Botta[b], Anna Sperotto[a], Lambert J.M. Nieuwenhuis[a], Mattijs Jonker[a], Abhishta Abhishta[a]

[a] *University of Twente, Drienerlolaan 5, Enschede, 7522 NB, The Netherlands*
[b] *University of Napoli Federico II, Corso Umberto I 40, Napoli, 80138, Italy*

Abstract

DDoS attacks could have an economic motive such as extortion, but also social and political such as hacktivism and state-sponsored warfare. Hence, the monetary value of the targets does not always explain the victimization. To counter DDoS threats, cloud providers utilize more robust, distributed networks and implement DDoS evasion techniques such as Anycast. However, by hosting many targets under a single infrastructure, they also accumulate threats, which together might exceed the ability of the infrastructure.

We approach the victimization problem by using VIVA (value, inertia, visibility, and access) attributes of the DDoS victims to understand what makes them a suitable target for DDoS attacks. We conduct a large-scale analysis of DDoS attack incidents inferred from traffic data recorded by a network telescope over a five-year period. Using Alexa rank and content category of the domain names associated with the targeted IP addresses, we infer the targets' popularity and industry sector to estimate their VIVA

*Corresponding author
Email address: m.y.m.haq@utwente.nl (Muhammad Yasir Muzayan Haq)

attributes, especially, visibility and value. We reveal that more popular domain names suffer more DDoS attacks but during the COVID-19 pandemic, we observe more attacks targeting domain names regardless of their popularity. Our analysis of the top 100K most popular domain names shows that certain industry sectors have significantly higher threats of DDoS attacks and even repeat victimization. We also investigate the accumulated threats among cloud/data center providers and find that the proportion of customers from specific industry sectors statistically increases the threat of DDoS attacks for these providers significantly.
Keywords: DDoS, victimization, VIVA, popularity, industry sector, accumulated threat, cloud provider, customer portfolio

1. Introduction

In recent years, Distributed-Denial-of-Service (DDoS) attacks have become an imminent threat, since most critical processes rely on the availability of online systems. DDoS attacks overburden online systems with a huge amount of traffic or computational requests that would lead to resource exhaustion and ultimately, downtime. When the targeted systems serve highly critical functions, e.g., the domain name resolution, the impact of the downtime could be massive since it could create a ripple effect by also making other systems that rely on the target's services unreachable. For example, the infamous Mirai DDoS incident in 2016 targeting Dyn, a large managed DNS service provider, crippled the service for about two hours leaving many users incapable of accessing some of the largest sites, including Paypal, Twitter, and Amazon (Abhishta et al., 2018).

In addition to its massive impact, motivation for DDoS attacks cannot be fully explained based on financial benefits to the attacker as they do not always result in direct monetary compensation. Political reasons, such as cyber warfare, and socio-cultural causes, such as revenge and hacktivism, could also motivate a DDoS attack (Abhishta et al., 2020). For example, the DDoS attacks on Dyn in 2016 were believed to be initiated by "angry gamers" trying to offense Sony PlayStation Network[1]. Another large DDoS incident in 2015 targeted GitHub, a popular online code management platform, and more specifically the URLs of two projects that attempted to overcome Chinese state censorship "The Great Firewall of China". Many speculated that the Chinese government launched the attacks to force GitHub into terminating the projects (Goodin, 2015). Killnet, a pro-Russian hacktivist group, threatened to launch DDoS attacks on the SWIFT network to cripple the Western banking system in protest to U.S. involvement in the Russia-Ukraine conflict (IBM X-Force, 2023). Threat actors could also use DDoS attacks for extortion: they demand a ransom for terminating the ongoing attacks, such as the incident that targeted a European gambling company in February 2021 (Ilascu, 2021).

Organizations must implement a proper strategy to manage the risk of DDoS attacks that threaten their system. In this paper, we empirically investigate the threat of DDoS attacks on different business sectors. Based on Routine Activity Theory (RAT) in criminology we hypothesize that the perceived value and visibility of an organisation's IT resources influence their

[1] `https://www.justice.gov/opa/pr/individual-pleads-guilty-participating-internet-things-cyberattack-2016`

threat of DDoS attacks (Yar, 2005; Abhishta et al., 2020). We have two main goals. *First*, we aim to assist organizations in managing DDoS risks, specifically, to estimate the threat of DDoS attacks against their networks. *Second*, we seek to facilitate cloud/data center providers in measuring the level of DDoS threat against their networks based on their customers' business sectors. To achieve these objectives, we characterize victimization in DDoS attack incidents to answer two main research questions:

RQ 1 How do organizational characteristics that represent their value and visibility attributes, namely, popularity and industry sector, correlate with the threat of DDoS attacks?

RQ 2 How do the variety of business sectors catered by a data center affect the threat of DDoS attacks?

Despite a large number of incidents, most of DDoS attacks are not publicly disclosed by the victims for various reasons including avoiding any reputation damage and the insignificant impact of the incident[2,3]. Hence, we rely on Internet measurement data to infer the incidents from network traffic activities. We analyze metadata of DDoS attack incidents derived from traffic data recorded by a network telescope for five years between August 1st, 2016, and July 31st, 2021. We elaborate further about the datasets in Section 2.1.

The contributions of this paper are as follows:

[2]`https://www.ncsc.gov.uk/blog-post/why-more-transparency-around-cyber-attacks-is-a-good-thing-for-everyone`

[3]`https://www.cnbc.com/id/100491610`

- We perform a large-scale study on DDoS attack incidents to reveal how the popularity and industry sector of the targets affect victimization.
- We reveal industry sectors with higher levels of DDoS threats.
- We demonstrate that the customer portfolio of cloud/data center providers has an impact on the threat of DDoS on these providers.
- We present quantitative results on DDoS threats that could help organizations and cloud providers estimate their threat of DDoS in case of change in customer portfolio (i.e. acquisition of new customers from a specific industry sector).

1.1. Background

In this section, we briefly discuss the definitions and theories that guide our analysis and explanation of the results. Risk and threat are two concepts with several definitions depending on the context and the subject domain in which they are used. In this work, we refer to Gordon-Loeb model (Gordon and Loeb, 2002) that proposes risk of attack to information set as expected loss L using Equation (1), where

- v is *vulnerability* or the probability of success if such an attack happens,
- t is *threat* or the probability of an attack happening to the asset, and
- λ is *potential loss* or the value of the information set in the absence of any security measure.

$$L = v \times t \times \lambda \tag{1}$$

In this work, we focus on quantifying the threat t or the probability of a DDoS attack happening to a network that hosts an online system. For a more accurate measurement of threat, we provide a more nuanced understanding of how DDoS threat is dependent on the target characteristics, such as popularity and industry sector, rather than using the probability of such an attack to any target. To understand the nature of target selection in DDoS attacks, we adopt certain theories from criminology including RAT and VIVA frameworks.

Routine Activity Theory (RAT) states that a crime happens when there is a suitable target, a motivated offender, and no capable guardian present (Cohen and Felson, 1979). To explain a suitable target, VIVA (Value, Inertia, Visibility, and Access) (Yar, 2005) refers to four attributes that determine the likelihood of a victim being targeted, namely, *value* of the victim, *inertia* or physical obstacles of the victim, *visibility* of the victim, and *access* or reachability of the victim. As with any other criminals, we may assume these theories also apply to DDoS victimization. In this study, we propose to use *popularity* and *industry sector* of a target as means to estimate its visibility and value attributes respectively.

Industry sector of an online system may reflect the perceived value and influence motive of the offender (Abhishta et al., 2020). Certain industry sectors might reflect a higher monetary value, e.g., bank and financial institutions, or a higher political value, e.g., social media and news portal, than other sectors. Hence, the industry sector may greatly affect the motive, e.g., hacktivists or state-sponsored offenders are more likely to target governmental bodies than other organizations (ENISA, 2023; Abhishta et al.,

2020).

To counter DDoS threats, cloud/data center providers offer a solution by hosting the systems on more robust, larger, and distributed infrastructures capable of withstanding a massive traffic or computation load. However, cloud services also introduce some issues. Infrastructure sharing, which is one of the underlying principles of cloud services, hides the actual target of DDoS attacks on the cloud provider's network, hindering the ability to analyze the victimization, i.e., the target selection. For example, an IP address in a web hosting provider might be shared with up to thousands of customers' websites and when DDoS attacks hit the IP address, we could not pinpoint the actual target among these customers. Moreover, by hosting more organizations in a single network, cloud providers might attract more frequent and concentrated DDoS attacks to the network. For example, threat actors that seek to incapacitate the financial system in a country would launch DDoS attacks on the infrastructure that serves most banks to optimize the impact. A previous investigation also revealed that the Internet nowadays is already centralized; dominated by a handful of giant cloud providers (Kashaf et al., 2020). By serving a huge number of customers, successful DDoS attacks on their infrastructures would create a massive impact by also disrupting the systems of thousands or millions of their customers. In this study, we provide a nuanced understanding of the collective threat of DDoS attacks on these data center/cloud infrastructures.

2. Material and methods

Here, we explain the data and the step-by-step process in this study. We refer to *dedicated IP addresses* as IP addresses that are only associated with a single organization, while *shared IP addresses* are shared between multiple organizations. Inferring the actual targets of the attacks on shared IP addresses is highly sophisticated if not impossible. Therefore, we only focus on attacks on dedicated IP addresses which we can identify and confirm the actual targets.

2.1. Data Sources

We use multiple data sources along with additional supplementary data as summarized in Figure 1 that illustrates which source provides which data (white arrows), and how these data correlate to each other (black arrows). To infer the popularity of the targets and the industry sectors of the targets, we use the domain names associated with the target IP addresses.

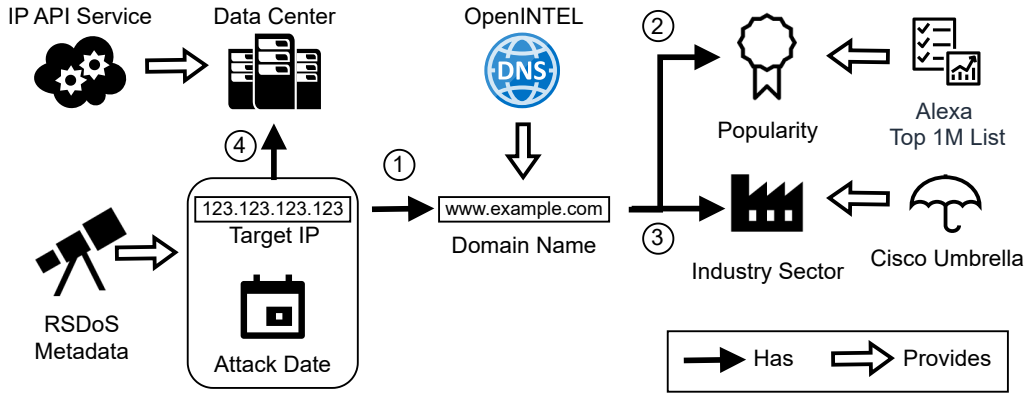


Figure 1: Dataset development flowchart

Denial-of-Service Attack Data. CAIDA UCSD Network Telescope collects and analyzes data on randomly and uniformly spoofed Denial-of-Service (RDoS) attacks by monitoring unsolicited IPv4 traffic (CAIDA, 2021). The CAIDA UCSD Network Telescope passively observes traffic directed at a large, unassigned block of IP addresses. From this traffic, backscatter packets are extracted, which are responses sent by victims of RDoS attacks, enabling researchers to infer attack incidents. The dataset includes information such as the target IP address, statistical details about the attack (e.g., duration and packet count), geolocation data of the target, and Autonomous System (AS) information. In this study, we use metadata of DoS incidents collected for five years since August 1st, 2016 up to July 31st, 2021. We select this time frame to include periods before and after COVID-19 started expecting to observe any changes the pandemic has brought to DDoS threats.

Passive DNS Measurement. OpenINTEL is a large-scale, active DNS measurements platform, covering >65% of the global DNS name space (van Rijswijk-Deij et al., 2016). Relevant to our analysis, we extract from the OpenINTEL dataset the domain names associated with the target IP addresses (①).

Top Domain List. Alexa Top 1 million list ranks the most popular domains based on traffic metrics derived from user interactions such as unique visitors and page views. These metrics are primarily collected from users who have installed the Alexa Traffic Rank browser extension (Le Pochat et al., 2019). For our study, we downloaded the Alexa top 1 million dated 2016-08-01, coinciding with the start date of the RSDoS dataset (Scheitle et al., 2018). We extracted the first 100K domains from this list, representing a section

of the Internet's most popular sites. We also use the rank from Alexa to indicate the popularity of each domain name (②).

Domain Categorization. Cisco Umbrella Investigate API provides a comprehensive overview of domains, including details such as IP addresses and Autonomous System Number (ASN) information. We use the Cisco Umbrella Investigate API (Cisco Systems, 2024b) to extract the category of Alexa domain names (③). Specifically, we leverage its "Domain Status and Categorization" feature for this purpose (Cisco Systems, 2024a).

IP Address Information. IP API Service[4] provides meta-information for both IPv4 and IPv6 addresses (ipapi.is, 2023). In this work, we used this service to obtain IP WHOIS data, including organization name, domain name, prefix, and prefix assignee. Also, the API enabled us to identify whether an IP address was associated with a data center, providing insights into the company and network from the prefix WHOIS information (④).

2.2. Data Analysis

Our data analysis consists of three main phases: target identification, target characterization, and threat analysis. Figure 2 summarizes the overall process of our analyses.

Target Identification. The first phase is identifying the targeted organization in DDoS attacks. In our dataset, each attack is associated with a targeted IP address. However, an IP address can be shared with multiple organizations, for example, in shared web hostings such as GoDaddy. Therefore, we first

[4] `https://ipapi.is`

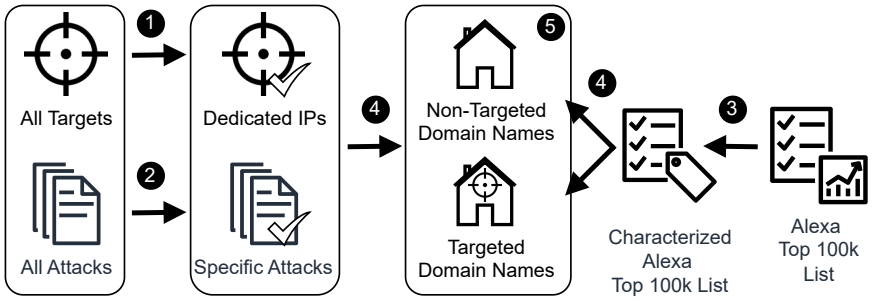


Figure 2: Data analysis flowchart.

need to identify IP addresses that belong to a single organization. To perform this task, we utilize historical DNS records collected daily by OpenINTEL for domain names under three large gTLDs, i.e., `com`, `net`, and `org`, and all new gTLDs under ICANN's Centralized Zone Data Service (CZDS)[5]. The steps, which are represented by ❶ in Figure 2, are the following:

Step 1 We retrieve `A` records for all target IP addresses on 2024-04-08 to check which IP addresses are still hosting a domain name nowadays. We exclude IP addresses with no associated domain names.

Step 2 We identify the date of the first attack suffered by each remaining IP address and then retrieve `A` records of the IP address on the date of the first attack.

Step 3 We extract the second level domain (SLD) of all domain names associated with the IP addresses on the first attack incident targeting it and on the more recent day. Then, we count the unique SLDs associated with each IP address and exclude addresses with more than one

[5] `https://newgtldprogram.icann.org`

unique SLD. We refer to the remaining IP addresses as dedicated IP addresses.

Target Characterization. The second phase involves obtaining more information about the targets which includes the popularity rank, domain category, and data center identity of the targets identified from the previous phase. To query the popularity and domain category, we need the domain names of the targets instead of the IP addresses. The steps, which are represented by ❸ in Figure 2, include:

Step 1 We retrieve domain popularity rankings from Alexa's top 1 million domain list on the first date in the dataset, i.e., 2016-08-01, obtained from the TU Munich repository[6] (Scheitle et al., 2018).

Step 2 We query the popularity rank of all domain names associated with the dedicated IP addresses if they belong in the top 1 million list. Otherwise, no popularity rank is given.

Step 3 We obtain the content category tags of all domain names associated with the top 100k Alexa list from Cisco Umbrella Investigate. Note that one domain name could have multiple tags or none.

Step 4 We use the IP API service to query the identity of the data center that hosts the dedicated IP addresses. If the IP address does not belong to any known data center, no information is returned.

[6] `https://toplists.net.in.tum.de/archive/alexa/`

Threat Analysis. The next phase is analyzing the DDoS threat across different characteristics as follows:

Step 1 We filter DDoS incidents in our dataset that target the dedicated IP addresses identified from the previous steps as illustrated by ❷ in Figure 2.

Step 2 We identify domain names in Alexa's top 100k list that have been targeted by DDoS attacks, as depicted by ❹ in Figure 2. At the end of this step, we have the final dataset (❺) that contains all domain names in Alexa's top 100k list including their industry sectors, popularity ranks, data centers, and summary of DDoS attack incidents targeting them.

Step 3 We use dataset ❺ to analyze the influence of popularity and industry sector on DDoS victimization.

Step 4 Finally, we investigate the influence of customer portfolio on the level of DDoS threat against cloud providers by aggregating the DDoS attacks targeting the customers of each data center provider.

3. Results

We applied our methodology to discover attacks with identifiable targets from all DDoS attacks in the dataset. Table 1 summarizes the comparison between all attacks and the targeted ones within the 5 years. We labeled more than 34K out of 2.8M unique IP addresses as dedicated IP addresses. These 34K IP addresses are associated with more than 35K unique domain

names since one IP address might be associated with multiple domain names. However, there are only 34K unique SLDs out of these 35K domain names reflecting the presence of different TLDs for the same SLD.

	#Attack	#Unique IP	#Unique Domain	#Unique SLD
All	16,478,338	2,820,583		
With identifiable target	110,278	34,768	35,644	34,235

Table 1: Comparison between the initial DDoS incident dataset and the one with identifiable targets.

We also identified 110K attacks to these 34K dedicated IP addresses out of 16M attacks recorded by the network telescope during the 5 years. Figure 3 shows the cumulative number of attacks with identifiable targets compared to all attacks. We observe a significant and sudden increase in attacks in early 2017. Our further analysis revealed that this increase is related to an anomaly, which we will elaborate on in the following section.

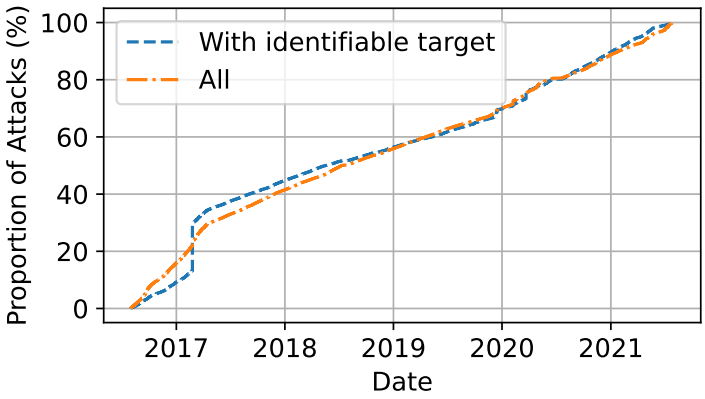
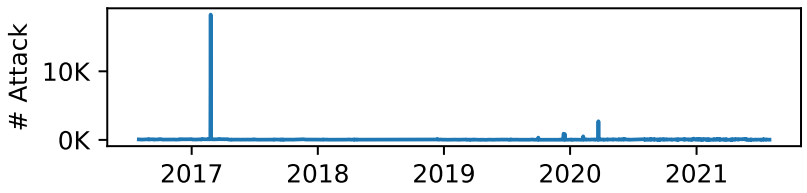
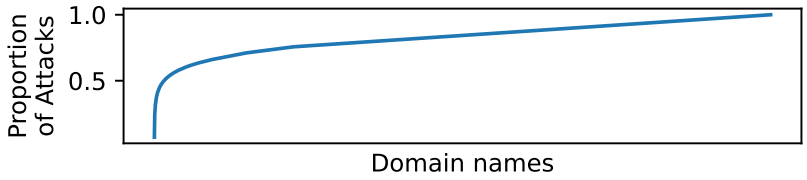


Figure 3: Cumulative plots of attacks with identifiable targets over time compared to all attacks.

Anomalies. We observe two extreme outliers in our dataset as illustrated in Figure 4. *First*, on February 25$^{\text{th}}$, 2017, we observe a huge number of attacks targeting almost the entire IPv4 space. On this day alone, there were more than 18K DDoS attack incidents which account for ∼20% of all attack incidents in the whole dataset outnumbering any other day as shown in Figure 4a. *Second*, there is a single domain name, i.e., `guff.com`, which suffered DDoS attacks consistently, every day, for the entire time frame. The total number of attacks targeting this domain name is more than 8K attacks outnumbering all other domains as shown in Figure 4b. Since we focus on the victimization patterns, we exclude data related to these two outliers, i.e., attack incidents on February 25$^{\text{th}}$, 2017, and attacks targeting `guff.com`



(a)



(b)

Figure 4: Outliers in the number of attacks by (a) attack date, and (b) target domain names. Plot (a) depicts the longitudinal number of attacks recorded per day. Plot (b) illustrates the cumulative proportion of attacks with the x-axis representing domain names sorted by the largest number of suffered attacks.

in the following analyses to avoid bias. Further investigation is required to discover the actual reasons behind these data outliers. However, we might speculate that the first anomaly is related to a large attack campaign with *random targets*.

Key takeaway: *There are two extremes in the dataset: attacks targeting almost all IPv4 addresses on a specific day, and attacks happening every day targeting a specific domain name. These observations suggest that the RAT theory is not always applicable to DDoS attacks, i.e., the targets are chosen regardless of their characteristics (suitability).*

3.1. Popularity of Targets

More popular websites are exposed to more DDoS attacks considering several aspects that make them a suitable target (see VIVA in Section 1). *First*, popular websites are not only well known to legitimate Internet users but also to malicious actors, hence, more likely to be victimized due to **high visibility**. *Second*, they have a relatively **high value**, especially, websites that gain their revenue from traffic such as online news and advertising. Malicious attackers with financial motivation are more likely to victimize these websites expecting to gain higher monetary incentives. Here we aim to validate these propositions by measuring how DDoS threats vary across websites with different popularity.

Figure 5 illustrates the number of attacks targeting websites with different popularity ranks. Darker areas indicate a higher number of attacks. From the plot, we can observe that the attacks are more concentrated on the more popular websites corroborating our belief that more popular websites attract

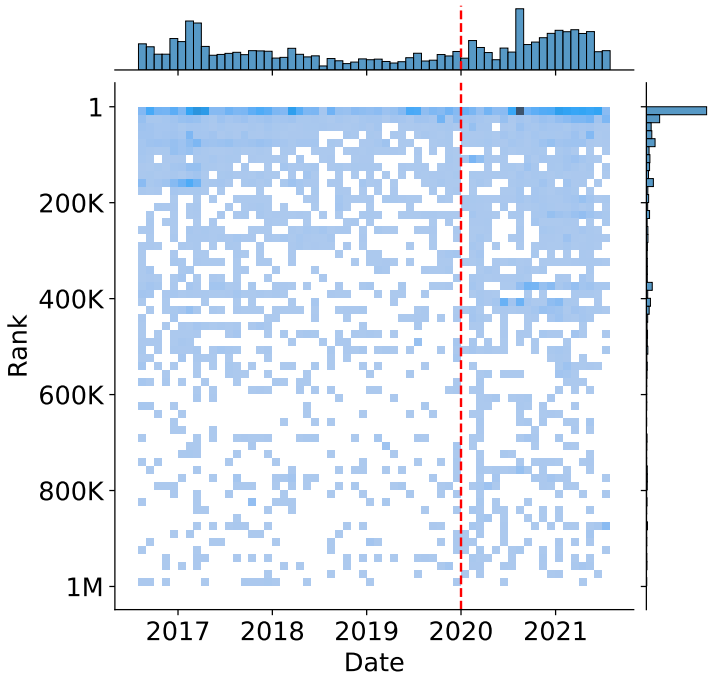


Figure 5: Number of attacks by popularity rank. Darker areas indicate more attacks. The red dashed line indicates the beginning of the COVID-19 pandemic.

more DDoS attacks. This observation is rather consistent over time, but we can observe a slight change of trend since early 2020, namely, less popular websites got attacked more frequently than before. This phenomenon might be related to the COVID-19 pandemic which also started in the beginning of 2020. During the pandemic, most people must stay at home and work from there which increased the use and reliance on the Internet and most online systems, which in the end motivates malicious attackers to attack any websites (Ganti and Yoachimik, 2020; Yoachimik and Singh, 2020).

***Key takeaway**: More popular websites suffered more DDoS attacks compared to less popular websites. During the COVID-19 pandemic, we observe an increasing number of DDoS attacks which also target less popular websites.*

3.2. Industry Sector of Targets

In addition to popularity, the industry sector of the targets might also influence the DDoS threat. *First*, the industry sector of a target might indicate its monetary value. For example, websites of financial institutions such as banks have relatively higher monetary value considering the amount of money they exchange daily. In addition, online banking websites also store personal and sensitive information of their users that are highly attractive to malicious actors. *Second*, certain industry sectors are more dependent on the Internet and online systems. For example, e-commerce companies, such as Amazon and AliExpress, mainly rely on online purchases made on their platforms as the primary sources of revenue. Hence, they need to maintain the availability of their websites. Otherwise, every minute of downtime that their websites suffer would incur a significant loss. *Third*, DDoS attacks can also be politically motivated. Hence, certain categories of website, such as government agency websites, might attract more DDoS attacks initiated by its political rivals or hacktivists.

Here, we analyze how DDoS threats vary across different industry sectors reflected by the category of domain hosted in the targeted IP address. However, to minimize bias, we limit the target to websites in the top 100k based on their Alexa rank. Then, we normalize the number of attacks targeting websites in a certain category with the total number of top 100k websites in that category. Hence, a high number of attacks on a category will be justified if the number of websites in the category is also high.

3.2.1. Industry Sectors of Top 100k Domain Names

We use 100k of the most popular domain names for DDoS threat analysis as they represent the most important ones in the entire Internet ecosystem. Therefore, by analyzing DDoS threats targeting these domain names, we hope to infer conclusions that are also applicable to other domain names. To provide context, we first identify the category of these top 100k domain names regardless of being the target of DDoS attacks in our dataset. Figure 6 shows the top industry sectors of the top 100k most popular domain names. The right-hand side plot indicates the proportion of the top 100k domain names that belong in each category. The left-hand side plot indicates the popularity rank of domain names in each category proportionally in equal-sized ranges. There are a small percentage of domain names (7%) that do not have any content category from Cisco Umbrella which we label as *unknown*.

Although most categories are evenly distributed across popularity ranges, websites in certain categories, such as News/Media, Search Engines, and Streaming Video, are more concentrated in the upper popularity range. In other words, many websites in these categories are very popular and consumed worldwide, such as CNN, Google, and YouTube. This might also indicate that global popularity is an important aspect for businesses in these industry sectors.

Key takeaway: *Business Services, Software/Technology, and News/Media are the top 3 categories by size which account for >20% of the top 100k most popular domain names. Domain names in the News/Media and Search Engines categories are concentrated in the highest popularity range.*

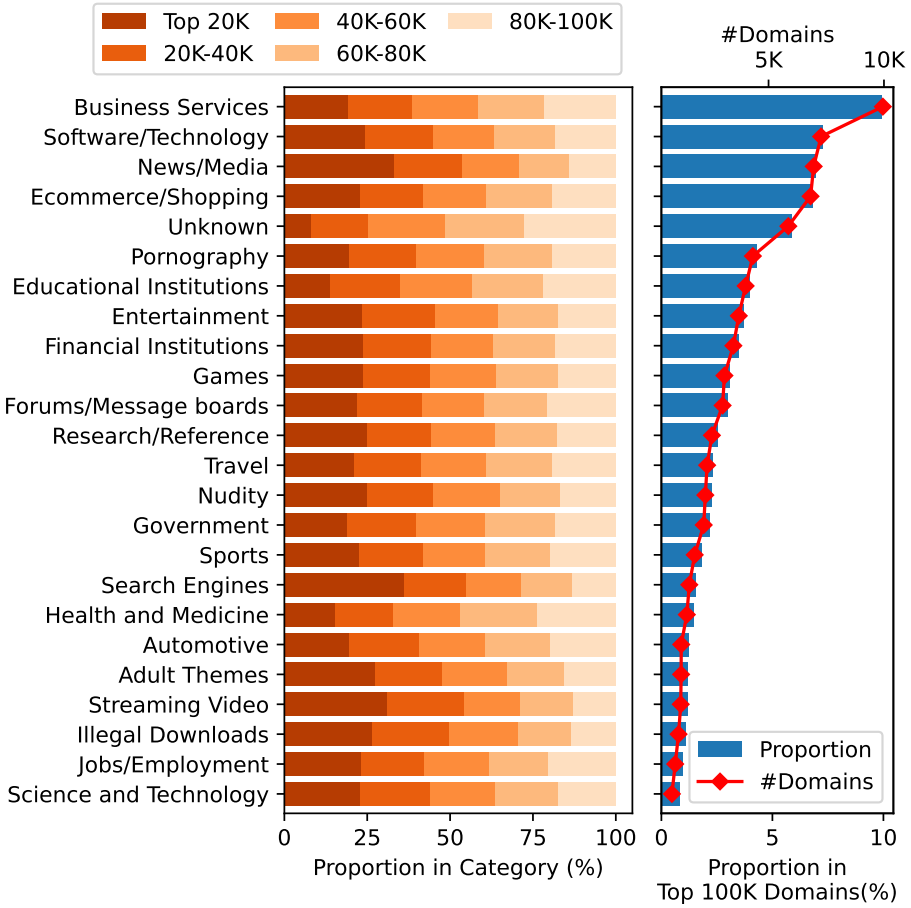


Figure 6: Composition of domain category groups among top 100k domain names by their popularity rank.

3.2.2. DDoS Threats in Various Industry Sectors

Figure 7 summarizes the attacks targeting the top 100k domain names grouped by domain categories from Cisco Umbrella Investigate content categorization. We provide three measurements in Figure 7, namely, the number of top 100k domain names in a category, the proportion of domains in each category that were targeted by DDoS attack, and how many times the same domain names were attacked. Number of domain names per category indicates the population size estimate of the category. For example, News/Media

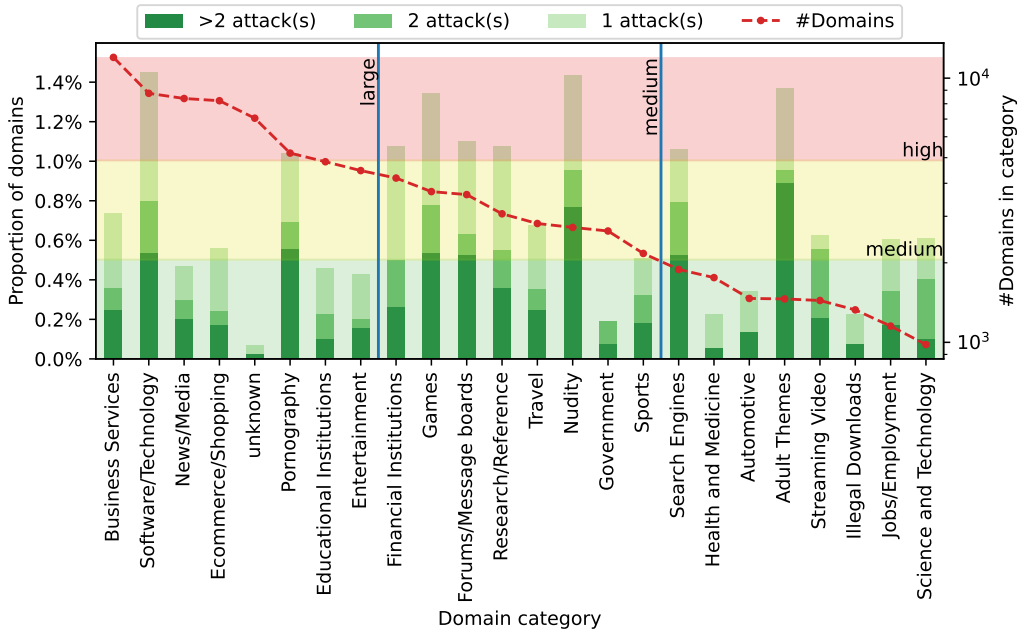


Figure 7: Top 100k domains targeted by DDoS attacks in the most frequent categories, characterized by the number of attacks each domain name suffers. Green, yellow, and red shaded areas indicate low, medium, and high threat levels.

is the third largest category with more than 8k domain names. Despite that, the number of News/Media domain names that were attacked is relatively low which is less than 0.5% of them. Meanwhile, the Software/Technology category with the second most number of domain names got attacked relatively more often than News/Media, indicating a higher threat of DDoS attacks targeting this category. Using the first two metrics, i.e., the size and the threat, we may classify the categories into different groups. If we split the measurement of each metric into three levels equally, we can create a two-dimensional matrix and classify the categories as shown in Table 2.

From Table 2, we can estimate the severity of DDoS threat to these categories. If two categories have similar levels of DDoS threat, then the

Threat Level	Population Size		
	Large *(L)*	Medium *(M)*	Small *(S)*
High *(3)*	Software/Tech. Pornography	Financial Inst. Games Forums/Message boards Research/ Reference Nudity	Search Engines Adult Themes
Medium *(2)*	Business Services Ecommerce/ Shopping	Travel Sports	Streaming Video Jobs/Employment Science & Tech.
Low *(1)*	News/Media Educational Inst. Entertainment	Government	Health & Medicine Automotive Illegal Downloads

Table 2: Classification of domain categories based on population size and DDoS threat level.

category with a larger population is more at risk. Recall the Gordon-Loeb formula of risk which is a product of the vulnerability, the threat, and the value of the asset. Hence, the accumulated risk of a DDoS attack on a domain category increases with more online systems in the category because the accumulated value also increases assuming that every online system has a certain valuation. Therefore, we may conclude that Software/Technology is the category with the highest risk of DDoS attack. According to the documentation of Cisco Umbrella, this category includes "*sites about computing, hardware, and technology, including news, information, code, and vendor information.*" (Cisco Systems, 2024c). Hence, DDoS attacks on domain names in this category might target the customers of a cloud computing service provider. Another category with similar traits (*L-3*) includes Pornography.

Pornography websites have been frequently targeted by hacktivists who are against the business or by cybercriminals aiming to steal sensitive user data such as the case of Ashley Madison data breach (Cox, 2016). We can also observe a high level of threat in categories with similar audiences such as Nudity and Adult Themes.

Despite having less number of domain names, other categories with high levels of DDoS threats (*M-3* and *S-3*) should also anticipate a large number of DDoS attacks on their systems, especially, websites of Financial Institutions which possess confidential data as well as play a crucial role in the e-commerce ecosystem. Websites such as Paypal and Stripe provide payment services which are essential for e-commerce websites. Online gaming websites might be targeted by their players such as the infamous DDoS attack in 2016 targeting the PlayStation website that also took down many other websites served by the same managed DNS provider, Dyn (Abhishta et al., 2018).

Categories with medium threat levels (*L-2, M-2,* and *S-2*) still require to anticipate a considerable amount of DDoS attacks. The availability of websites in these categories is highly important for businesses. Websites in certain categories such as E-commerce and Travel, e.g., ticketing websites, gain most of their revenue from transactions made on their websites, therefore, any downtime, including those caused by DDoS attacks, will lead to direct financial losses. To provide quality services to the users, Streaming Video, Sports, and Jobs/Employment websites have to keep their system reliable and available all the time especially if they charge their users for a subscription fee. Downtime of their services might lead to long-term losses if the users are disappointed and then decide to unsubscribe following the

incident (Haq et al., 2022). Meanwhile, according to the documentation (Cisco Systems, 2024c), domain names in the Business Service category include "*sites for corporations and businesses of all sizes, especially company websites*". Therefore, the attacks could be orchestrated by the target's competitors attempting to disrupt the business.

Additionally, websites in categories with low threat levels (*L-1* and *M-1*) are not totally safe from DDoS threats. Despite having a lower threat level, domain names in the News/Media category should still implement measures to mitigate DDoS threats considering that their businesses are dependent on their websites' availability. These websites are highly reliant on advertising fees and sometimes subscription fees to run their businesses. Hence, every minute of downtime might incur significant losses, especially for the big ones such as `nytimes.com` and `theguardian.com`. Most customers of Dyn in this category reacted to the downtime incident in 2016 by using secondary DNS providers indicating the importance of keeping their websites online (Haq et al., 2022). Moreover, websites of Governments and Educational Institutions should remain cautious of DDoS threats considering their possession of highly sensitive information. Government websites have been frequently targeted by politically motivated DDoS attacks, often between nations, as a part of cyber warfare (Abhishta et al., 2020). Online entertainment websites should also maintain their service quality and availability to avoid customer dissatisfaction.

We should highlight that certain categories with low threat levels, e.g., Illegal Downloads, have a rather small number of domain names (*S-1*). Therefore, the threat levels might be less representative than the larger categories.

Further studies to investigate DDoS threats on websites in these categories are required. Figure 7 also shows how many websites in each category that suffer DDoS attacks more than once. We discuss this further in the following section.

Key takeaway: *A large number of domain names in a category does not necessarily increase the number of DDoS targets in the category. Certain categories get attacked relatively more often than others. Software/Technology is the category with the highest risk considering the large population and the high level of DDoS threats of the category.*

3.2.3. Repeat Victimization

Repeat victimization in criminology refers to "*repeated offenses against the same person, household, business, or other target however defined*" (Fisher and Lab, 2010). In the context of DDoS, multiple attacks on the same website are more likely than other types of cyber crimes such as data breaches. DDoS attacks seek to take down the target systems by exhausting the resources but the impact of successful attacks is not permanent. Once the systems are back online, the attackers can launch another DDoS attack to the systems to cause another downtime. Ultimately, the attackers might also ask for a ransom, otherwise, the attackers will keep attacking the systems disrupting the victims' business for a longer period. Therefore, it is crucial to investigate and measure the threat. Here, we analyze the number of DDoS attacks targeting the same domain name as an indication of repeat victimization.

Figure 8 presents a heatmap illustrating the number of targeted domain names in the top categories characterized by the number of attacks suffered.

To facilitate better understanding, we classify the attack frequency into three groups, i.e., 1 attack, 2 attacks, and more than 2 attacks, assuming that attacks on the same target for more than twice is a good indication of repeat victimization. From Figure 8, we observe that repeat victimization happens

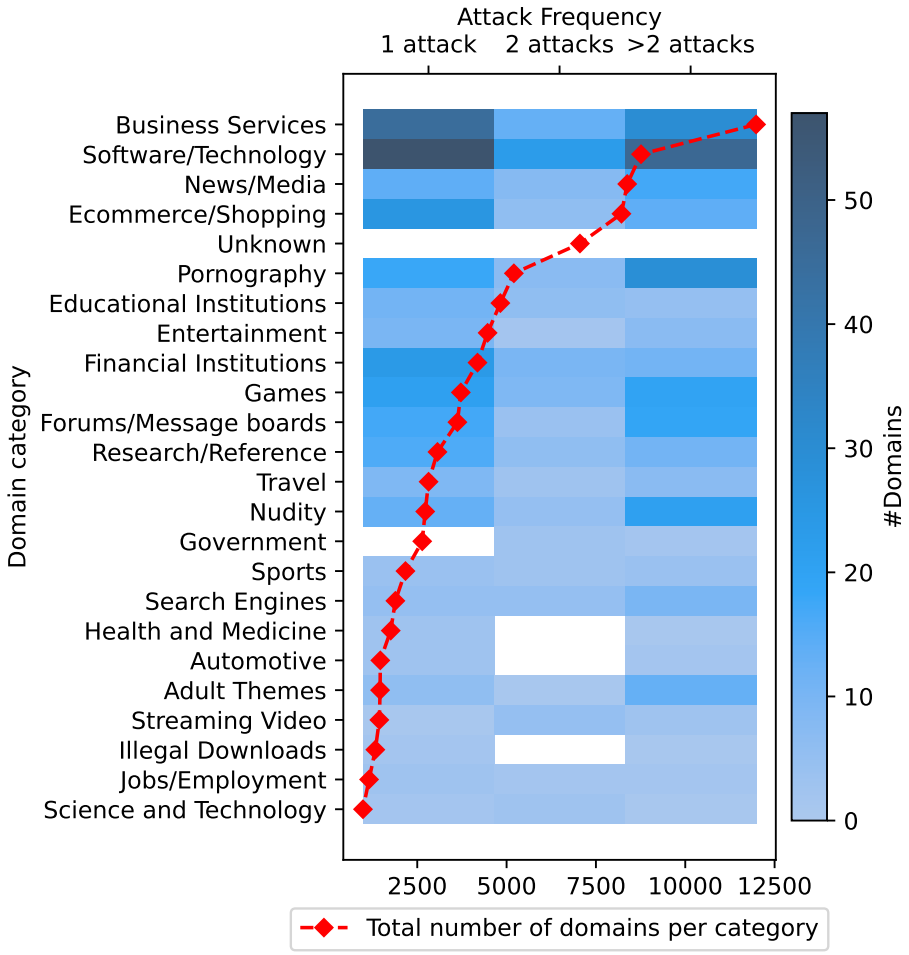


Figure 8: Heatmap representing the number of websites by attack frequency per domain category. Darker cells indicate more domains in the category targeted with a certain frequency of attack.

more often among websites in the Pornography, Nudity, and Adult Themes categories which also have similarities in their audiences and the type of content that they serve. In these categories, the threat of getting multiple DDoS attacks are relatively higher since more websites suffered multiple DDoS attacks than the websites that only experienced it once. Meanwhile, in the Business Services, Software/Technology, and Ecommerce/Shopping categories, majority of the websites only experienced the attack once.

Key takeaway: *Certain categories, including Pornography, appear to experience repeated DDoS attacks. However, without data on the identities of the attackers or the specific consequences of these attacks, it is not possible to conclude whether these domains are targeted more frequently by the same actors or whether the impact of these attacks is more severe. The observed pattern may suggest a higher susceptibility to being targeted, but the exact nature and implications of this repeat victimization remain unclear.*

3.2.4. Statistical Analysis

We conduct a statistical analysis to further evaluate how the industry sector of a domain name influences its likelihood of getting a DDoS attack. For this analysis, we continue using the top 100k most popular domain names as our sample. We define a dependent variable $is\_attacked$ to indicate if at least one IP address associated with a domain name has been targeted by a DDoS attack, i.e., $is\_attacked = 1$ if the domain name has been targeted and $is\_attacked = 0$ if otherwise. Moreover, we transform the category labels given to each domain name into one-hot encoded variables. For example, if `example.com` has two category labels, e.g., Travel and Television, then both

Travel and *Television* variables will be equal to 1 while the other variables will be equal to 0. We only include categories with at least 100 domain names and we exclude all attacks that happened on February 25^{th}, 2017, and targeting on `guff.com` since we consider them as outliers as discussed in the beginning of Section 3.

We use a Generalized Linear Model (GLM) from `statsmodels` package (Perktold et al., 2024) in Python considering its flexibility to accommodate different distributions of the dependent variable and link functions between the independent and dependent variables. Since *is_attacked* is a binary variable, we perform logistic regression by using GLM with binomial family and logit link function. Table 3 presents the regression coefficients of the top categories.

The results in Table 3 indicate that some categories have a significant correlation with DDoS victimization. Websites in categories with positive correlation, such as Software/Technology, Financial Institutions, and Games, have a higher probability of becoming a DDoS target than websites in categories with negative correlation, such as News/Media, Entertainment, and Government. P values indicate how statistically significant the result is, e.g., $p < 0.05$ means that the result is statistically significant with a 95% confidence interval.

Compared to the other categories, Government websites have a lower probability of becoming a DDoS target. This result does not necessarily reject the fact that many DDoS attacks are politically motivated. It only indicates that government websites are less likely to be a DDoS target compared to other websites such as Games and Financial Institutions. In other

words, politically motivated attacks are statistically less than attacks that are motivated by other factors such as financial.

Domain Category	n	Coefficient		Std. Error
Business Services	11,984	-0.023		(0.131)
Software/Technology	8,762	0.614	**	(0.119)
News/Media	8,376	-0.611	**	(0.184)
Ecommerce/Shopping	8,214	-0.206		(0.174)
Pornography	5,204	0.094		(0.243)
Educational Institutions	4,826	-0.279		(0.235)
Entertainment	4,469	-0.509	*	(0.255)
Financial Institutions	4,186	0.468	**	(0.18)
Games	3,720	0.602	**	(0.169)
Forums/Message boards	3,624	0.311		(0.18)
Research/Reference	3,066	0.439	*	(0.191)
Travel	2,817	0.034		(0.251)
Nudity	2,723	0.407		(0.286)
Government	2,639	-1.233	**	(0.458)
Sports	2,170	-0.258		(0.317)
Search Engines	1,888	0.222		(0.245)
Health and Medicine	1,758	-0.914		(0.529)
Automotive	1,466	-0.626		(0.458)
Adult Themes	1,460	-0.012		(0.294)
Streaming Video	1,440	-0.323		(0.367)
Illegal Downloads	1,327	-1.271	*	(0.589)
Jobs/Employment	1,152	-0.089		(0.393)
Science and Technology	981	-0.201		(0.418)

*$p < 0.05$; **$p < 0.01$

Table 3: Regressions of DDoS victimization on domain categories.

3.2.5. Reflection

We have shown that the popularity and industry sector of domain names can help understand victimization in DDoS attacks. These factors estimate the threat by indicating the **value** and **visibility** of domain names as targets. However, our study lacks a solid view of **inertia** and **access** attributes. Further research is needed to analyze how these remaining attributes affect DDoS victimization. In addition, while the data suggest a pattern in the types of targets chosen for DDoS attacks, the underlying motivations of attackers remain speculative. However, determining the actual motives behind DDoS attacks at a large scale is extremely difficult, as attackers often employ DDoS-on-Demand providers, such as booters, to carry out these attacks on their behalf. This use of third-party services acts as a proxy, masking the true identity and intent of the original instigators. Furthermore, the nature of DDoS attacks—particularly those utilizing reflection and distributed methods—further complicates attribution by obscuring the command and control center. These layers of indirection and distribution make it nearly impossible to trace the origin or motivation of the attack with certainty.

3.3. Customer Portfolio and DDoS Threats on Data Center Providers

We have observed that domain names in certain categories have relatively higher DDoS threats than others. Our next question is: *what would happen if many websites in the categories with high DDoS threats are hosted on the same infrastructure?* In this section, we approach the question by analyzing the data centers of these websites and the number of DDoS attacks that are accumulated to each data center provider.

From the IP API service, we query the data center information of all IP

addresses used by the top 100k most popular domain names. Data centers in this study were broadly defined to include hosting providers, CDNs, and cybersecurity services because these entities often share similar infrastructure and face comparable risks in the context of DDoS attacks. This approach allowed for a comprehensive analysis of DDoS vulnerabilities across a wide range of services. We obtain two data that indicate the provider of the data center, namely, the name such as 'Google' and the domain name of the provider such as `www.google.com`.

3.3.1. Clustering Data Centers

Since the names of the data centers are not always consistent even though they refer to the same provider, such as Google or Google LLC, we do some text mining on both the data center name and domain name to assist us in clustering data centers from the same providers together.

First, we perform data cleansing including lowercasing, stopwords removal, and word deduplication. Before removing the stopwords, we replace all punctuation symbols with a space. The stopwords that we remove from the text include 1) single-letter words, 2) common TLDs and subdomains, e.g., `com`, `net`, `org`, and `www`, 3) common terms used in the company names, e.g., 'hosting', 'technologies', and 'technology', and 4) abbreviations of the corporation type, e.g., 'gmbh', 'llc', 'inc', and 'ltd'. We replace duplicated words to minimize the noise since we concatenate both names and domain names as a single text, e.g., 'google llc www.google.com', so that the text ultimately would transform into 'google' after applying all data cleansing tech-

niques. *Second*, we vectorize the texts using TF-IDF[7] from `sklearn` package in Python to transform the texts into their numeric vector representations (scikit-learn developers, 2024b). *Third*, we use DBSCAN[8] from the same Python package to group similar vectors in the same clusters (scikit-learn developers, 2024a). We choose DBSCAN since it does not need a predefined number of clusters (providers). Instead, the algorithm groups the text vectors based on their Euclidian distances to each other. To optimize the clustering algorithm, we first transform the vectors using `StandardScaler` from the same Python package to normalize the values into a normal distribution, i.e., its standard deviation is equal to 1.

3.3.2. Data Center Providers of The Top 100k Domain Names

Before we analyze the DDoS threat on the data center providers, we first look at the market share of each provider. Among the top 100k domain names, 29% of them do not use any data centers according to our dataset. Figure 9 illustrates the market share of the top 20 providers against the popularity ranks of their customers in the top 100k domain names. Note that a domain name might use multiple data center providers. Cloudflare and Amazon AWS dominate the market share by hosting around 12% and 8% of the top 100k domain names. Customers of CloudFlare and DigitalOcean are more concentrated among the more popular ones while the customers of Amazon AWS are evenly distributed in all popularity ranges. Another large provider, i.e., OVH, indicates a concentration of customers in the lower

[7]Term Frequency × Inverse Document Frequency

[8]with $eps = 0.75$ and $min\_samples = 1$

popularity ranks.

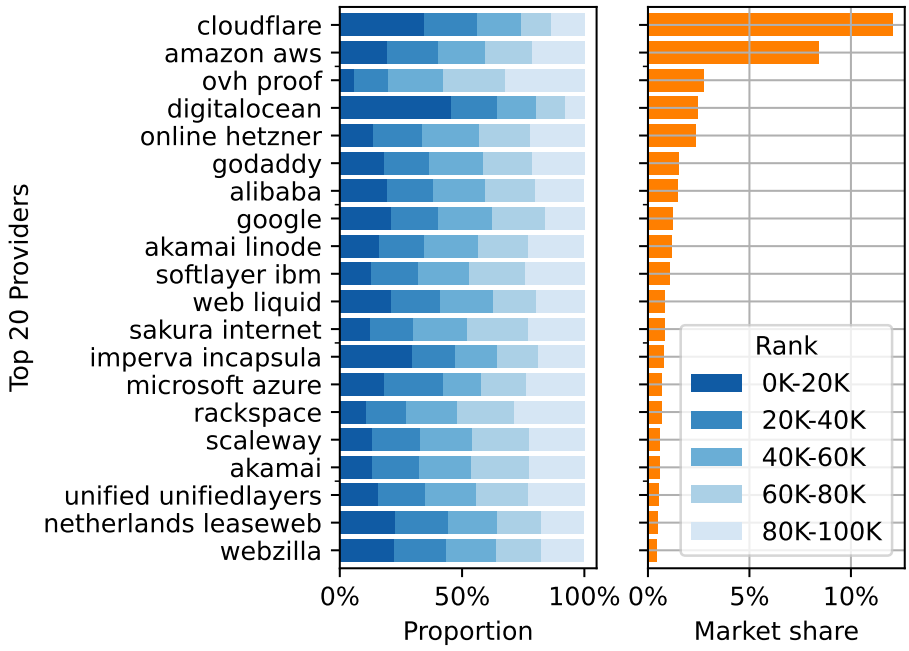


Figure 9: Top 20 data center providers of top 100k domain names across popularity ranks.

When we analyze the market share of the providers by industry sectors of their customers, we can observe the composition of customers of each data center provider as shown in Figure 10. For example, Cloudflare's customers as the largest provider are mostly domain names in News/Media, Entertainment, and Pornography categories. Meanwhile, customers of Amazon AWS, the second largest provider, are dominated by domain names in Business Services, Software/Technology, and E-commerce categories.

Despite a high number of domain names in Educational Institutions and Government categories, only a few use third-party data centers. These domain names most probably host their systems on private infrastructures that

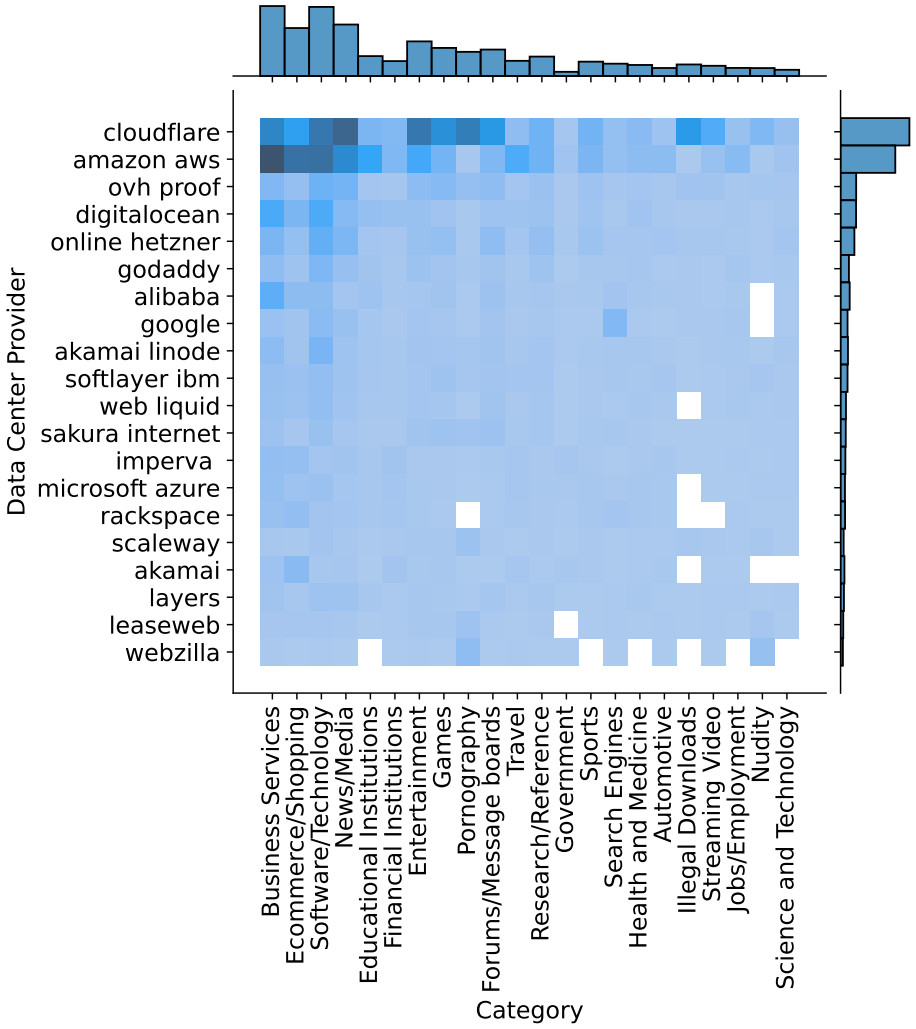


Figure 10: Top 20 data center providers of top 100k domain names across industry sectors (top 20).

they own and manage themselves. They might also use *community clouds* which are infrastructures that are shared among other similar institutions,

such as SURF[9] and Logius[10] in the Netherlands for Dutch educational institutions and government agencies, respectively. Meanwhile, domain names in Financial Institutions that host their system on third-party data centers are also limited. Since financial industry is highly regulated by the governments, the regulations might also dictate how financial institutions should act when using outsourced cloud infrastructure, especially to host sensitive financial data, for both security and privacy reasons (European Union Commision, 2016). However, further investigation is required to confirm these speculations.

Key takeaway: *Despite the customers of most data center providers are rather evenly distributed across all popularity ranges, customers of certain providers, such as Cloudflare and DigitalOcean, are concentrated in the highest popularity range while OVH's customers are dominated by less popular domain names.*

3.3.3. DDoS Attacks on Data Center Providers

We follow up our discussions on the market share and customer portfolio of data center providers by analyzing how the composition of customers from different industry sectors correlate with the DDoS attacks that a provider suffers. For this analysis, we first aggregate all attacks targeting IP addresses that belong to the same data center providers. Then we group the customers of each provider according to the DDoS threat level classification as presented in Table 2. We assume that other smaller categories that are not present in

[9]https://www.surf.nl/

[10]https://www.logius.nl/

Table 2 to have a low threat level. We associate the risk of customer with the threat level of its domain category, e.g., we refer to high risk customers as the ones that belong to domain categories with a high level of threat. Finally, for each provider, we calculate the proportion of each type of customers, i.e., high, medium, or low risk customers.

Figure 11 summarizes the results of our analysis. The first plot (left) presents the proportion of customers with different levels of DDoS threat, while the second plot (right) compares the number of customers and the collective number of DDoS attacks per each data center provider. In general, we may observe from the second plot a positive correlation between the estimated size (the number of customers) and the total number of DDoS attacks suffered by a provider, i.e., larger providers get more attacks, such as Cloudflare, DigitalOcean, GoDaddy, and Alibaba. However, there exists some variation in the total number of attacks per provider that the number of customers does not explain. For example, OVH suffers similar number of DDoS attacks with Cloudflare even though OVH's customers are only less than a third of Cloudflare's customers. Another example is Amazon AWS which suffers significantly less attacks than Cloudflare even though the numbers of their customers do not vary that much.

We hypothesize that the number of customers from categories with a high level of DDoS threat can explain these variations. To test our hypothesis, we conduct a regression analysis with percentages of each threat level as the independent variables and the total number of attacks per provider as the dependent variable.

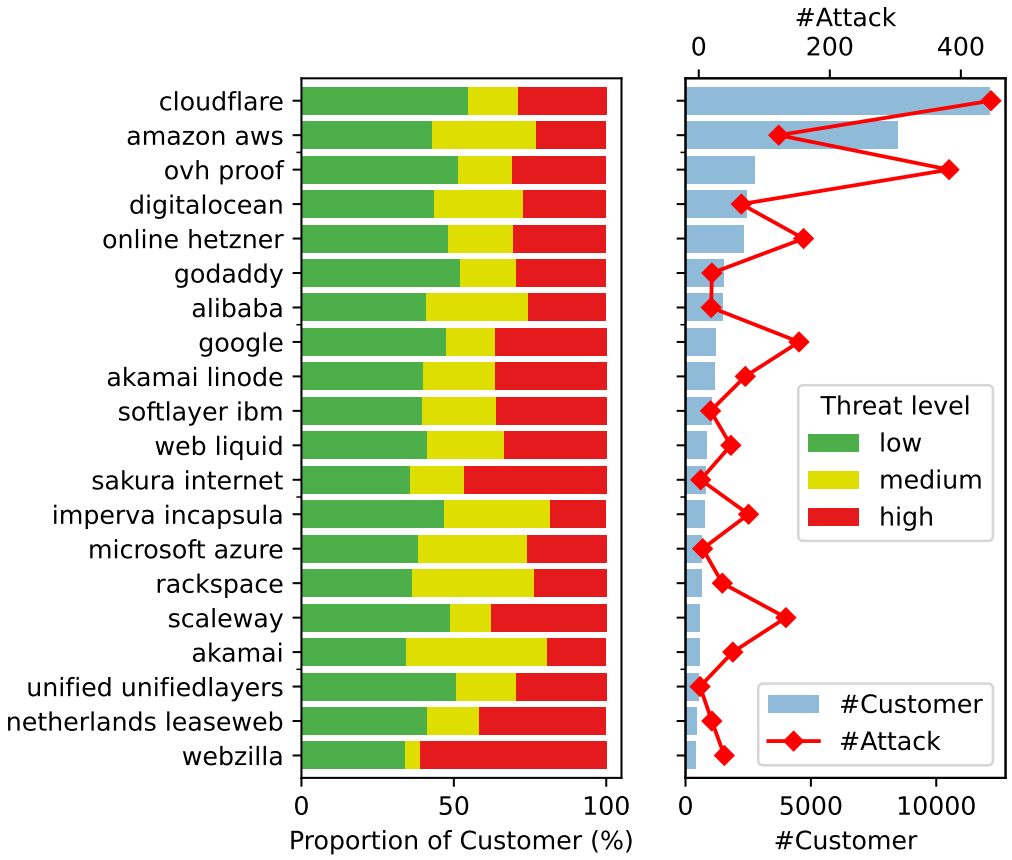


Figure 11: Proportion of customers by DDoS threat levels per data center provider (left) and comparison between the number of customers and the collective number of attacks per data center provider (right).

3.3.4. Statistical Analysis

According to our initial analysis on the dependent variable, i.e., total number of attacks per provider, the values fit closest to a Gamma distribution. Hence, for our regression analysis, we use GLM from `statsmodels` package in Python with Gamma family distribution and logit link function. Table 4 presents the results of the regression analysis. We exclude providers with fewer than 20 customers as well as those with zero attack. The results show that the percentages of high- and medium-risk customers positively

correlate with the number of attacks targeting the providers. Meanwhile, the percentage of low risk customers negatively correlate with the number of attacks. In addition, the result shows that the positive correlation of the percentage of high-risk customers with the number of attacks is statistically significant with 99% confidence interval.

Variable	Coefficient	Std. Error
Percentage of high risk customers	124.612 **	(47.07)
Percentage of medium risk customers	-52.977	(52.02)
Percentage of low risk customers	120.000	(88.619)

$^{**}p < 0.01$

Table 4: Regressions of number of attacks per provider on the percentage of customers with different risk levels.

Key takeaway: *In addition to the number of customers, the proportion of customers of data center providers also positively correlate with the number of attacks on the providers.*

4. Discussion

We have presented the results of various analyses that seek to reduce the complexity of understanding the threat of DDoS attacks. We believe that these results would benefit the stakeholders especially to improve their strategy to manage DDoS risks.

Implications for Organizations. The results of this study could assist decision-makers in organizations to make well-informed decisions regarding their risk management process. In general, they could prepare a better strategy against

DDoS attacks especially if their industry sectors have a high level of DDoS threat. They could use our results in the risk assessment process in which they need to identify and quantify the risk of DDoS attacks to plan a proper mitigation strategy. Using the results from Section 3.2, they could adjust the likelihood factor according to their industry sector instead of using a generic probability of DDoS attacks targeting their networks. Furthermore, they could use the risk calculation to justify the security investments for DDoS protection and mitigation. When it comes to cloud services, organizations should be aware that some providers may cater to a higher proportion of high-risk customers, potentially attracting more DDoS attacks. However, these providers might also be better equipped to defend against such threats, effectively minimizing disruption despite facing a higher volume of attacks. It is important to note, though, that this aspect was not examined in this study.

Implications for Network Service Providers. Network and cloud service providers, including web hosting, CDN, and DDoS protection services, can leverage this study's findings in several ways. First, they can estimate their DDoS risk based on their customer portfolio, justifying investments in infrastructure hardening if the risk is high. Second, providers can create a premium, more resilient network for high-risk customers, which can be offered at a higher price. Third, by implementing a KYC[11] process, providers can screen potential customers to estimate their DDoS threat level. This process allows providers to offer premium infrastructure to high-risk customers, reject

[11]Know Your Customers

them if necessary, or distribute them across multiple networks to manage risk effectively.

Implications for Academic Researchers. In this study, we demonstrate that DDoS threats might vary depending on several aspects including non-technical characteristics of the targets such as popularity and industry sectors. Therefore, it is crucial to include such non-technical aspects in the studies on DDoS risk as well as on other types of cybercrime. By not assuming the risks to be identical for all targets, we would produce conclusions that are closer to the real-world scenario, hence, create a more positive impact to society.

Centralization of Internet infrastructure. Kashaf et al. (2020) revealed that the Internet nowadays is highly dependent on a handful of giant cloud service providers due to the centralization of network infrastructures. Meanwhile, recent outages showed that the disruptions in large providers, such as Fastly in June 2021[12], Akamai in July 2021[13], and Amazon AWS in December 2021[14], could cause a massive impact that breaks the global Internet connectivity. Security companies and researchers could use our work in evaluating the collective risks of cloud providers, especially the large ones, against DDoS attacks by estimating the composition of their customers. If the majority of a provider's customers come from industry sectors facing a high level of DDoS threats, but the provider lacks proper infrastructure and adequate

[12] `https://www.fastly.com/blog/summary-of-june-8-outage/`

[13] `https://www.akamai.com/blog/news/akamai-summarizes-service-disruption-resolved`

[14] `https://www.cnbc.com/2021/12/09/how-the-aws-outage-wreaked-havoc-across-the-us.html`

protective measures, there is a high likelihood of significant disruptions due to DDoS attacks in the future. A large-scale analysis could help estimate the overall resilience of the Internet ecosystem, both regionally and globally, by assessing how well providers can mitigate these threats.

4.1. Related Works

In this study, we proposed an approach to understanding DDoS threats by analyzing the attack incidents, characterizing the targets, and exploring the collective risks among data center providers based on the number of high-risk customers. We took many inspirations from previous studies on similar topics.

Several studies analyzed DDoS attack incidents for various objectives. Jonker et al. (2017) implemented the criteria from Moore et al. (2001) to infer randomly and uniformly spoofed DoS attacks (RSDoS) from traffic collected by the CAIDA network telescope and combined these inferences with reflection and amplification attacks learned from honeypot logs. Using the resulting comprehensive dataset of attacks of different types, Jonker et al. (2017) characterized attacks over time. Moreover, by combining attack target IP addresses with active DNS measurement data from the OpenINTEL project, Jonker et al. studied attacks on Web infrastructure and investigated if and when Web sites migrate to DDoS Protection Services after being attacked. Their work enabled many other studies to use the DDoS attacks data set and analyze DDoS attacks in a large scale, including our work. Sommese et al. (2022) also combined RSDoS data with OpenINTEL data, to investigate the impact of DDoS attacks on authoritative DNS infrastructure. They highlighted the amplified magnitude of impact if DDoS attacks tar-

get authoritative name servers that provide name translations for numerous domain names. They discovered the importance of anycast deployment in DNS infrastructure which is proven effective to minimize the impact of DDoS attacks.

Previous works have attempted to characterize DDoS attacks and victims to improve the counter-attack measures. Noroozian et al. (2016) profiled the victims of DDoS-as-a-service or the booters using the data from amplification DDoS honeypots. They discovered that most targets are users in access networks (broadband). Jonker et al. (2017) provided a framework to characterize DDoS attacks in a large scale, especially the technical characteristics of the attack methods and the targets. In this work, we focused more on the non-technical aspects of the targets aiming to assist organizations in making more strategic decisions. More recent work from Abhishta et al. (2020) explored the motivation behind DDoS target selections from the perspective of the attackers as criminal actors.

While cloud providers offer robust defenses against DDoS attacks, they are not immune. Abhishta et al. (2018) studied the impact of successful DDoS attacks on managed DNS service providers, focusing on two major 2016 incidents targeting NS1 and Dyn. These attacks disrupted DNS infrastructure, making many customer domain names inaccessible. They analyzed customer behavior post-incident, noting whether they stayed, used secondary providers, or left. Haq et al. (2022) extended this by examining non-technical factors affecting customer behavior after Dyn's downtime, highlighting the importance of availability in various sectors. While these studies examined DDoS impacts on providers, we focused on profiling threats to diverse cloud

consumers. Our results support Noroozian et al. (2016), who found that attacks on ISPs and web hosting networks increase with customer numbers. Similarly, we observed that high-risk customers elevate attack numbers on data center providers. To our knowledge, no prior study has explored the collective risk to cloud providers.

4.2. Limitations

We acknowledge some limitations in our work. *First*, our methodology to identify dedicated IP addresses based on the number of associated SLDs could have missed organizations that use single IP addresses to host multiple domain names. *Second,* since we do not use longitudinal data for target characteristics, i.e., popularity, industry sector, and data center provider, our analyses might overlook changes in these characteristics. A further investigation on a larger scale is required to accommodate these changes. We should also acknowledge that some data might be outdated and the patterns in DDoS victimization could have changed nowadays.

4.3. Future research

We still have a long way to go to fully understand the DDoS risks. However, we hope that our results could be a stepping stone towards it. Therefore, we call for a couple of research directions following this work to improve understanding of DDoS risks. We have presented characterizations of the DDoS targets to provide an initial understanding of the likelihood of becoming a DDoS target given the industry sector of an organization. However, we could improve the results by also characterizing the attacks. For example, we could analyze the time of attacks to reveal seasonal attack patterns per industry

sector (Abhishta et al., 2019). We could also study the volume and the duration of the attacks to observe if certain industry sectors suffer larger or longer DDoS attacks. Providing insight into these aspects will help network operators to plan better preventive measures, especially, when should they increase their capacity and by how much, to find the optimum strategy, i.e., the right balance between the performance and the cost.

Another direction for future research is to investigate the impact of DDoS attacks on the accessibility of targets with different characteristics. Not all DDoS attacks result in a total outage; some only cause performance degradation or have minimal effects. By studying these impacts, we can identify the industry sectors that suffer significant disruptions caused by successful DDoS attacks.

5. Conclusion

In this work, we address the challenges of managing DDoS risks by characterizing targets of DDoS attacks over 5 years. We analyze the popularity and industry sector of these targets, helping organizations estimate their level of DDoS threat based on similar characteristics. Our results confirm that on average more popular domains are more frequently attacked. However, we notice a trend change in the post-COVID-19 time period. Additionally, we identify specific industry sectors that are more often targeted repeatedly and significantly increase the likelihood of becoming a DDoS victim.

One of the common strategies to mitigate DDoS risks is to host online systems in a third-party data center equipped with a more robust and well-distributed network. We study the impact of the proportion of data center

customers from a given industry sector on the collective threat of DDoS attacks on that data center. Our results reveal that in addition to the number of customers, the proportion of customers from high-risk industry sectors (see Table 2) also significantly increases the number of DDoS attacks on a data center provider. In other words, data center providers that host a large proportion of high-risk customers are exposed to a higher DDoS threat. With increasing centralization in cloud services, it becomes more urgent to evaluate the DDoS threat levels among the providers to let them adjust their infrastructures' resilience accordingly.

CRediT authorship contribution statement

Muhammad Yasir Muzayan Haq: Conceptualization, Methodology, Investigation, Writing - Original Draft, Project Administration, Formal Analysis, Visualization. **Antonia Affinito:** Methodology, Investigation, Writing - Original Draft. **Alessio Botta:** Writing - Review & Editing, Resources. **Anna Sperotto:** Conceptualization, Writing - Review & Editing. **Lambert J.M. Nieuwenhuis:** Conceptualization, Supervision. **Mattijs Jonker:** Conceptualization, Resources, Writing - Review & Editing. **Abhishta Abhishta:** Conceptualization, Methodology, Writing - Review & Editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Aggregated daily RSDoS attack metadata is accessible for academic purposes upon request to CAIDA UC San Diego. OpenINTEL data is accessible for academic purposes upon request to the Design and Analysis of Communication Systems (DACS) group at the Faculty of Electrical Engineering, Maths and Computer Science (EEMCS), University of Twente. Cisco Umbrella and IP API data are accessible with commercial licenses. Alexa top 1 million list from TU Münich historical archive is publicly accessible.

Acknowledgments

We would like to thank Nikolai Tschacher for providing additional IP address information from `ipapi.is`. This work was supported by the Netherlands Organization for Scientific Research (De Nederlandse Organisatie voor Wetenschappelijk Onderzoek) under NWO:MASCOT project [CS.014], and Cisco Systems through the Sponsored Research Agreement "Research Project for Industry 4.0".

References

Abhishta, A., van Heeswijk, W., Junger, M., Nieuwenhuis, B., Joosten, R., 2020. Why would we get attacked? An analysis of attacker's aims behind DDoS attacks. Journal of Wireless Mobile Networks 11, 3–22. doi:10.22667/JOWUA.2020.06.30.003.

Abhishta, A., Junger, M., Joosten, R., Nieuwenhuis, L.J., 2019. Victim routine influences the number of ddos attacks: Evidence from dutch edu-

cational network, in: 2019 IEEE Security and Privacy Workshops (SPW), pp. 242–247. doi:10.1109/SPW.2019.00052.

Abhishta, A., Van Rijswijk-Deij, R., Nieuwenhuis, L., 2018. Measuring the impact of a successful ddos attack on the customer behaviour of managed dns service providers. Computer communication review 48, 70–76. doi:10.1145/3310165.3310175.

CAIDA, 2021. Aggregated Daily RSDoS Attack Metadata (Corsaro 2). https://catalog.caida.org/dataset/telescope_corsaro2_daily_rsdos. doi:https://catalog.caida.org/dataset/telescope_corsaro2_daily_rsdos. Dates used: 08/01/2016 to 07/31/2021. Accessed: 03/05/2024.

Cisco Systems, 2024a. Check status and categorization of domains - cloud security api - cisco devnet. https://developer.cisco.com/docs/cloud-security/check-status-and-categorization-of-domains/. (Accessed on 06/17/2024).

Cisco Systems, 2024b. Cisco umbrella investigate api: Domain status, risk score - cloud security api - cisco devnet. https://developer.cisco.com/docs/cloud-security/investigate-investigate-overview/. (Accessed on 06/17/2024).

Cisco Systems, 2024c. Legacy DNS Content Category Definitions. URL: https://docs.umbrella.com/umbrella-user-guide/docs/dns-content-category-settings. (Accessed on 06/13/2024).

Cohen, L.E., Felson, M., 1979. Social Change and Crime Rate Trends: A Routine Activity Approach. American Sociological Review 44, 588–608. URL: https://www.jstor.org/stable/2094589, doi:10.2307/2094589. publisher: [American Sociological Association, Sage Publications, Inc.].

Cox, J., 2016. As Hackers Continue to Target Porn Sites, Pornhub Launches Bug Bounty Program. URL: https://web.archive.org/web/20201108092100/https://www.vice.com/en/article/bmvvz4/pornhub-bug-bounty. (Accessed on 06/02/2024).

ENISA, 2023. Warfare and Geopolitics are Fuelling Denial-of-Service Attacks. URL: https://www.enisa.europa.eu/news/warfare-and-geopolitics-are-fuelling-denial-of-service-attacks. (Accessed on 06/19/2024).

European Union Commision, 2016. Commission Delegated Regulation (EU) 2017/565 of 25 April 2016 supplementing Directive 2014/65/EU of the European Parliament and of the Council as regards organisational requirements and operating conditions for investment firms and defined terms for the purposes of that Directive (Text with EEA relevance.). URL: http://data.europa.eu/eli/reg_del/2017/565/oj/eng. (Accessed on 06/17/2024).

Fisher, B., Lab, S., 2010. Encyclopedia of Victimology and Crime Prevention. Thousand Oaks, California. URL: https://sk.sagepub.com/reference/victimologyandcrime, doi:10.4135/9781412979993.

Ganti, V., Yoachimik, O., 2020. Network-Layer DDoS Attack Trends for

Q1 2020. URL: https://blog.cloudflare.com/network-layer-ddos-attack-trends-for-q1-2020. (Accessed on 06/17/2024).

Goodin, D., 2015. DDoS attacks that crippled GitHub linked to Great Firewall of China. URL: https://arstechnica.com/information-technology/2015/04/ddos-attacks-that-crippled-github-linked-to-great-firewall-of-china/. (Accessed on 06/18/2024).

Gordon, L.A., Loeb, M.P., 2002. The economics of information security investment. ACM Transactions on Information and System Security 5, 438–457. URL: https://doi.org/10.1145/581271.581274, doi:10.1145/581271.581274.

Haq, M.Y.M., Jonker, M., Van Rijswijk-Deij, R., Claffy, K., Nieuwenhuis, L.J., Abhishta, A., 2022. No Time for Downtime: Understanding Post-Attack Behaviors by Customers of Managed DNS Providers, in: 2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 322–331. doi:10.1109/EuroSPW55150.2022.00039. iSSN: 2768-0657.

IBM X-Force, 2023. Hacktivist Groups Threaten to Attack the SWIFT Network - Threat Activity IBM X-Force Report. URL: https://exchange.xforce.ibmcloud.com/threats/exchange.xforce.ibmcloud.com/threats/guid:6b1db065d2095671d07be6311da9910e. (Accessed on 06/18/2024).

Ilascu, I., 2021. 800Gbps DDoS extortion attack hits gambling company. URL: https://www.bleepingcomputer.com/news/security/800gbps-

ddos-extortion-attack-hits-gambling-company/. (Accessed on 06/18/2024).

ipapi.is, 2023. An Algorithm to Detect Hosting Providers and their IP Ranges. URL: https://ipapi.is/blog/detecting-hosting-providers.html. (Accessed on 04/28/2024).

Jonker, M., King, A., Krupp, J., Rossow, C., Sperotto, A., Dainotti, A., 2017. Millions of targets under attack: a macroscopic characterization of the DoS ecosystem, in: Proceedings of the 2017 Internet Measurement Conference, Association for Computing Machinery, New York, NY, USA. pp. 100–113. URL: http://doi.org/10.1145/3131365.3131383, doi:10.1145/3131365.3131383.

Kashaf, A., Sekar, V., Agarwal, Y., 2020. Analyzing Third Party Service Dependencies in Modern Web Services: Have We Learned from the Mirai-Dyn Incident?, in: Proceedings of the ACM Internet Measurement Conference, Association for Computing Machinery, New York, NY, USA. pp. 634–647. URL: http://doi.org/10.1145/3419394.3423664, doi:10.1145/3419394.3423664.

Le Pochat, V., Van Goethem, T., Tajalizadehkhoob, S., Joosen, W., 2019. TRANCO: A research-oriented top sites ranking hardened against manipulation, in: Network and Distributed Systems Security (NDSS) Symposium 2019. doi:10.14722/ndss.2019.23386.

Moore, D., Voelker, G.M., Savage, S., 2001. Inferring internet Denial-of-Service activity, in: 10th USENIX Security Sympo-

sium (USENIX Security 01), USENIX Association, Washington, D.C. URL: https://www.usenix.org/conference/10th-usenix-security-symposium/inferring-internet-denial-service-activity.

Noroozian, A., Korczyński, M., Gañan, C.H., Makita, D., Yoshioka, K., van Eeten, M., 2016. Who Gets the Boot? Analyzing Victimization by DDoS-as-a-Service, in: Monrose, F., Dacier, M., Blanc, G., Garcia-Alfaro, J. (Eds.), Research in Attacks, Intrusions, and Defenses, Springer International Publishing, Cham. pp. 368–389. doi:10.1007/978-3-319-45719-2_17.

Perktold, J., Seabold, S., Sheppard, K., ChadFulton, Shedden, K., jbrockmendel, j grana6, Quackenbush, P., Arel-Bundock, V., McKinney, W., Langmore, I., Baker, B., Gommers, R., yogabonito, s scherrer, Zhurko, Y., Brett, M., Giampieri, E., yl565, Millman, J., Hobson, P., Vincent, Roy, P., Augspurger, T., tvanzyl, alexbrc, Hartley, T., Perez, F., Tamiya, Y., Halchenko, Y., 2024. statsmodels/statsmodels: Release 0.14.2. URL: https://zenodo.org/doi/10.5281/zenodo.593847, doi:10.5281/ZENODO.593847. (Accessed on 06/17/2024).

van Rijswijk-Deij, R., Jonker, M., Sperotto, A., Pras, A., 2016. A High-Performance, Scalable Infrastructure for Large-Scale Active DNS Measurements. IEEE Journal on Selected Areas in Communications 34, 1877–1888. doi:10.1109/JSAC.2016.2558918.

Scheitle, Q., Hohlfeld, O., Gamba, J., Jelten, J., Zimmermann, T., Strowes, S.D., Vallina-Rodriguez, N., 2018. A Long Way to the Top: Significance,

Structure, and Stability of Internet Top Lists, in: Proceedings of the Internet Measurement Conference 2018, Association for Computing Machinery, New York, NY, USA. pp. 478–493. URL: https://dl.acm.org/doi/10.1145/3278532.3278574, doi:10.1145/3278532.3278574.

scikit-learn developers, 2024a. DBSCAN. URL: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html. (Accessed on 06/17/2024).

scikit-learn developers, 2024b. TfidfVectorizer. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html. (Accessed on 06/17/2024).

Sommese, R., Claffy, K., van Rijswijk-Deij, R., Chattopadhyay, A., Dainotti, A., Sperotto, A., Jonker, M., 2022. Investigating the impact of DDoS attacks on DNS infrastructure, in: Proceedings of the 22nd ACM Internet Measurement Conference, Association for Computing Machinery, New York, NY, USA. pp. 51–64. URL: https://dl.acm.org/doi/10.1145/3517745.3561458, doi:10.1145/3517745.3561458.

Yar, M., 2005. The novelty of 'cybercrime': An assessment in light of routine activity theory 2, 407–427. URL: https://doi.org/10.1177/147737080556056, doi:10.1177/147737080556056. publisher: SAGE Publications.

Yoachimik, O., Singh, A., 2020. Network-layer DDoS attack trends for Q2 2020. URL: https://blog.cloudflare.com/network-layer-ddos-attack-trends-for-q2-2020. (Accessed on 06/17/2024).

Curriculum vitae

Muhammad Yasir Muzayan Haq is PhD candidate at the Industrial Engineering and Business Information Systems (IEBIS) section at the University of Twente. His doctoral research focuses on measuring the risk in cloud outsourcing ecosystem to improve security decision-making. Prior to pursuing his PhD, he worked in an Indonesian start-up company as VP of Big Data Analysis.

Antonia Affinito is assistant professor at the Design and Analysis of Communication System (DACS) group at the University of Twente. Her research focuses on network monitoring, network security, and anomaly detection, with an emphasis on optimizing the energy consumption of Internet measurements and security practices.

Alessio Botta is associate professor at the Department of Electrical Engineering and Information Technlogies of the University of Napoli Federico II, Italy. He is also co-founder of NM2 a startup, spin-off of the University of Napoli Federico II, developing innovative products and services in the area of network monitoring. His research interests fall in the area of computer networks, and, in particular, of their performance measurement and improvement.

Anna Sperotto is professor at the Design and Analysis of Communication Systems (DACS) group at the University of Twente. Her research interests are network monitoring, intrusion detection, network security, and network management.

Lambert J.M. Nieuwenhuis is (emeritus) professor at the Department of High-Tech Business and Entrepreneurship (HBE) of the Faculty of Behav-

ioral, Management and Social sciences (BMS) of the University of Twente (UT). He is owner of the consultancy firm Knowledge for Business. His research interests include servitization in the Manufacturing Industry, Smart Industries, Cloud Computing and the economic impact of security threats through DDoS attacks.

Mattijs Jonker is assistant professor at University of Twente's Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS). His research is on network security in the broad sense and involves extensive data science and Internet measurement. He is also chief big data architect on the OpenINTEL project.

Abhishta Abhishta is assistant professor at the Industrial Engineering and Business Information Systems (IEBIS) section at University of Twente. His research focuses on empirically measuring the economic/financial impact of cyber attacks. In order to do so, he devices/adapts data-driven economic impact assessment techniques. He looks to help organisations make well-measured investments in security.