Scale matters: a Comparative Study of Datasets for DDoS Attack Detection in CSP Infrastructure

Clément Boin, Tristan Groléat, Xavier Guillaume, Gilles Grimaud, Michaël Hauspie

Scale matters: a Comparative Study of Datasets for DDoS Attack Detection in CSP Infrastructure

1st Clément Boin
OVHcloud
Roubaix, France
clement.boin@ovhcloud.com

3rd Tristan Groléat
OVHcloud
Brest, France
tristan.groleat@ovhcloud.com

3rd Xavier Guillaume
OVHcloud
Roubaix, France
xavier.guillaume@ovhcloud.com

5th Gilles Grimaud
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL
F-59000 Lille, France
gilles.grimaud@univ-lille.fr

6th Michaël Hauspie
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL
F-59000 Lille, France
michael.hauspie@univ-lille.fr

***Abstract*—Denial of Service (DoS) and Distributed Denial of Service (DDoS) are attacks designed to take down a service by exhausting its resources. Lots of research have been carried in the past decades to design efficient algorithms that can detect these attacks. However, most of the literature on DoS and DDoS detection consider the protection of a small or medium size businesses network. Usually, these networks consist in several workstations and servers protected by few firewalls that can analyze all incoming network traffic. So that the research on DoS and DDoS can be reproduced and analyzed, several datasets, reflecting this network infrastructures have been proposed in the literature. However, more and more businesses are migrating their services to the cloud and are renting servers from Cloud Service Providers (CSP). If the CSP wants to protect its customers from DoS and DDoS attacks, it must perform detection on its infrastructure. This kind of infrastructure is in no way comparable to the ones usually found in the literature. In this paper, we propose to compare publicly available state-of-the-art datasets with real network traffic captured on the infrastructure of a world-scale CSP and discuss their relevance in the context of detecting volumetric DDoS attacks on CSP infrastructure.**

***Index Terms*—DDoS, Cloud, Datasets, Hyperscalers**

I. Introduction

Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks are types of attacks aimed at making a service unavailable by depleting the resources of the target system. These attacks are used for various purposes such as financial blackmail, unfair competition or simply to cause damage. They have serious consequences for the victims, such as loss of income or damage to the reputation of the organization [1]. For example, Dyn's DDoS attack in 2016 was a massive attack that targeted the DNS system, causing major internet disruptions by making many websites inaccessible [2]. Since their appearance, these attacks have become more common and sophisticated, forcing organizations to protect themselves and their customers against them [3]. With our modern lifestyles becoming increasingly digitally-dependent, many organizations have embraced the use of Cloud Computing, which allows users to access computing resources via the Internet, rather than operating their infrastructure to provide services to their customers. This paradigm offers several advantages, such as flexibility, scalability, availability, security and reduced costs compared to purchasing and managing their infrastructure. Computing resources in the Cloud Computing model are provided by Cloud Service Providers (CSPs) [4]. This new way of producing and consuming IT infrastructure makes CSPs a target for attackers as they host numerous services making them a lucrative target [1]. In light of the escalating magnitude of DDoS attacks and the costs involved, it is impractical for CSPs to consistently over-provision their infrastructures as a defense mechanism. Furthermore, the strategy of continually subjecting clients to mitigation measures is ineffective due to the introduction of persistent latency and the inherent risks of false positives. It is in this context that researchers are working to create DDoS attack detection and mitigation systems that can quickly identify and respond to attacks while minimizing their impact. To propose detection and mitigation solutions, researchers must validate their experiments [5]. In addition to determine the limitations and performance of proposed solutions, validation is needed to convince stakeholders of the effectiveness of detection and mitigation of attacks. Indeed, the implementation of security measures against these attacks represents a significant investment [6]. As we will see in more details, there are several techniques for validating detection and mitigation solutions, but the datasets used by the community – whether they come from live traffic captures, simulated or emulated environments – present several limitations. Datasets suffer from the complexity of the simulation, the differences between test environments and real environments, the amount of data required for testing, as well as the high risks and costs associated with testing [7], [8]. In addition, constraints related to sensitive data present in network traffic captures often make these datasets unavailable or limited to the lab environments, not allowing the reproduction of infrastructures such as those of CSPs [9].

Therefore, in this study, we show that, while being valuable assets, the dataset currently provided by the literature are not representative of the trafic that can be observed at a CSP level. Moreover, we demonstrate through measurements and analysis

that one of the main issue when trying to detect volumetric DDoS attacks for a CSP is the relatively small amount of trafic reprensented by those attacks with regards to the total capacity of the CSP network infrastructure.

II. Problem statement

To design solutions for detecting and mitigating DDoS attacks, researchers need datasets representative of the type of infrastructure they seek to protect. As explained in Behal *et al.* [6], most of the solutions proposed in the state-of-the-art to validate DDoS detection and mitigation solutions are based on simulations, emulations, tests in real conditions and on the analyzes of publicly available real datasets. All these methods of generation have their advantages and limitations. They observe that the use of publicly available traces – whether they come from a laboratory environment or a real infrastructure – is increasingly widespread. In their article Ring *et al.* [7], review several datasets some of which contain commonly used DDoS attack scenarios, including the KDD Cup 1999, DARPA 1998, NSL-KDD, UNSW-NB15, and CICIDS2017 datasets. For each dataset, the authors present its characteristics, limitations and advantages as a resource for intrusion detection research. They also discuss how each dataset was collected, processed, and labeled, as well as the types of attacks represented therein. They end with a discussion on the limitations of existing datasets and future challenges for network-based intrusion detection research. They emphasize that existing datasets do not cover all possible situations, and that it is important to continue developing new datasets to reflect real threats.

In addition, the datasets found in the literature do not always represent the characteristics of a CSP. First of all, CSPs have a highly geo-distributed infrastructure, which means that incoming traffic can come from multiple points of presence (PoPs) located in different geographical locations. This reality is not reflected in the datasets we studied. Another aspect missing is the ratio between legitimate and attack traffic. Most of the datasets, when they do not only contain attack traffic, are far from the ratios found on the infrastructure of a CSP. For example, Microsoft which in 2021 mitigated a DDoS attack of more than 3.5Tbps [10]. This attack represents only a tiny part of the network capacity of Microsoft Azure, since it was already announced in 2017 “Within a given region, we can support up to 1.6Pbps of inter-datacenter bandwidth.” [11], which gives an idea of the total capacity of their infrastructure.

Furthermore, in the context of a CSP, DDoS attacks can target the CSP itself, as well as customers of its infrastructure. This duality implies that in addition to knowing the characteristics of its traffic, the teams in charge of the security of the CSP must know the characteristics of the traffic of each of its customers, which is not always possible. Although some CSPs only offer Software a Service (SaaS) services, most offer a large service catalog, such as Platform as a Service (PaaS) or Infrastructure as a Service (IaaS). For the latter, the customer is often free to configure their services, and for the specific case of IaaS, the customer may be administrator on these machines, so they can change the service topology at any time [12].

In this article we propose to study the statistical characteristics of a selection of datasets used in the literature to compare these same characteristics on the real production traffic observed on the backbone of our infrastructure as a world wide CSP. The expected outcome of our study is to assess whether publicly available datasets are suitable for research on DDoS detection at the scale of a large CSP.

III. Related work

In the literature, several papers aim to study the relevance and quality of datasets publicly available to researchers.

First, Thomas *et al.* [13] analyze the DARPA dataset. They review the characteristics of the dataset, which is one of the most used to evaluate detection systems. They conclude that due to its size, variety and representativeness of network attacks, it has its place among the datasets used but they also point out that the dataset has certain limitations, including weaknesses in the representation of some types of attacks and biases in the distribution of attacks. They draw the reader’s attention to the use of multiple datasets to evaluate intrusion detection systems in order to reduce bias and improve the representativeness of attacks. Similarly, Dhanabal *et al.* [14] also propose to review a dataset, NSL-KDD, which is a subset of the KDD Cup 1999 dataset also widely used by the community. In addition to this review, they propose several algorithms of classifications and conclude that NSL-KDD is a useful tool for the evaluation of intrusion detection systems based on classification algorithms. Malowidzki *et al.* [8] show the importance of choosing an appropriate dataset to experiment with a detection solution, as well as the importance of using publicly available datasets. They explain the advantages and limitations as well as the collection methodology and then the pre-processing applied to the data. The authors do not offer a solution to compare the datasets to the production traffic of a CSP. Their contribution focuses on exposing the problem of the lack of datasets in our field.

Bhuyan *et al.* [15] review the various methods, systems and tools used for the detection of anomalies in computer networks. In the section on the evaluation criteria, they review several commonly used datasets and come to the conclusion that although the datasets, which they define as benchmarks, are very useful, they are not representative of real traffic. The authors did not study if those datasets reflect a CSP environment.

Gharib *et al.* [16], evaluate several datasets to determine their quality by measuring diversity, size, normalization, relevance, credibility and availability. Their work also includes analysis and tools to generate synthetic data sets for intrusion detection. They show that most available datasets lack diversity, standardization and credibility, and do not accurately represent real data from computer networks. In this study, the authors do not seek to compare the datasets to real traffic nor do they propose a dataset that meets the criteria they propose.

Ring *et al.* [7], review several widely used datasets. For each of them, they present the technical characteristics, the types of intrusions simulated, the performance measures evaluated and their strengths and weaknesses. They also evaluate their ability to simulate real attack scenarios, their size, their diversity and

their representativeness. The authors outline the challenges in creating new datasets for network-based intrusion detection, such as the difficulty of simulating realistic attack scenarios and the need to protect private data.

Cordero *et al.* [17] provide a review of datasets used in the field of Intrusion Detection Systems (IDS) up until 2018. Furthermore, they propose comparing the datasets based on measures such as the number of packets and the variation in entropy of different fields in the IP packet header, which we have also chosen to replicate in our study. Additionally, the authors propose a dataset generator that takes a pcap file of background traffic as input, along with a description of the attacks the user wishes to inject, in order to generate datasets that adhere to the statistical properties of the original background traffic.

Damasevicius *et al.* [18] propose a new dataset based on a real university network. They also compare it to several datasets from the literature that we are studying. The authors' work focuses on an infrastructure that is approaching the size of a CSP. Additionally, they employ statistical methods to compare and analyze their traffic, which aligns with the methods we also utilize in our comparison. Based on the results they have obtained, we can conclude that despite the size of their infrastructure, the impact of volumetric DDoS attacks on statistical measures remains significant, unlike in a CSP infrastructure, as we demonstrate in the subsequent part of our study.

Maciá-Fernández *et al.* [19] provide a literature review of some of the datasets that we have also selected in our study. Following this literature review, they describe how they constructed their datasets using captured data from a Tier 3 ISP portion. They detail the evolution of traffic in response to the attacks they encountered using the same kind of statistical measures as those employed in our study. It is noteworthy that the dimensions of their infrastructure are similar to those of a CSP.

In the literature that we have reviewed, none seek to compare the datasets to the traffic that a CSP can observe on its infrastructure. The authors seek either to compare the datasets between them, to identify which is the most relevant to evaluate a given detection method, or they establish a list of criteria that a dataset must respect to be considered as of good quality and evaluate the dataset against these criteria.

IV. Datasets selection

As noted by both Behal *et al.* [6] and Camargo *et al.* [20], most of the publicly available datasets – such as DARPA 1999 [21] or CAIDA 2007 [22] – are either outdated or lack the features essential for developing detection tools. Until now, we are not aware of a comparative study of the usability of the datasets in the literature to detect DDoS attacks against CSP infrastructures. This is why we have selected datasets that seem the closest to a CSP, in particular those that contain both legitimate traffic and attacks, include at least several hours of capture, with several volumetric DDoS attacks. Moreover, we only study publicly available datasets. Ring *et al.* [7] compare 34 datasets commonly used in the field of anomaly-based network intrusion detection systems. From their review, we can keep a list of 16 datasets that seem useful for detecting and mitigating DDoS attacks at first sight.

The table I presents a comprehensive overview of the key attributes pertaining to the various datasets employed in the scholarly literature. It furnishes pertinent details encompassing the year of publication, cumulative size, duration, data format, public accessibility, and the specific types of legitimate traffic and attacks addressed within each respective dataset. By conducting this comparative analysis, we can ascertain the datasets that will be employed for comparison against our production traffic in our study.

We have chosen the following datasets for our study: CICIDS 2017 [35], ISCX 2012 [36], UNSW-NB15 [37], CSE-CIC-IDS2018 [35], DDoS 2019 [38] URG'16 [19] and, LITNET-2020 [18], as these datasets are publicly available, contain both legitimate and attack traffic captures, and importantly span a sufficiently long time interval to allow for the evolution of legitimate traffic patterns in response to seasonal effects. Additionally, we have selected datasets in Packet format (PCAP) or Netflow format to ensure that all metadata utilized by the existing literature can be extracted.

In the remainder of this article we will compare the selected datasets against each other, using the characteristics selected in the datasets with what we observe in a CSP infrastructure.

V. Metrics

We have selected a series of statistical characteristics among the most used in the literature to describe the traffic present in the datasets and to compare them with each other, as well as with the production traffic of the CSP we study [18], [39]. We computed these characteristics by aggregating the packets over one-minute time window. We have chosen the following statistical characteristics to describe the traffic:

- *Number of source IP addresses per destination IP address*: This feature helps describing the complexity of traffic flows and can potentially indicate the presence of DDoS attacks utilizing botnets. For each one-minute time window, we calculate the maximum number of unique source IP addresses that have communicated with a single destination IP address.
- *Total number of packets observed*: This metric helps describing the overall traffic volume and can potentially indicate the presence of volumetric attacks. For each one-minute time window, we record the total number of packets observed.
- *The distribution of protocols*: This metric helps describing the diversity of protocols present in the traffic and can potentially indicate the presence of attacks targeting specific protocols. For each one-minute time window, we compute the ratio between the various protocol counters (TCP, UDP, ICMP, or OTHER) to determine the protocol distribution. To calculate the "Protocol distribution" metric, we first need to compute the protocol counts and ratios for each protocol in the one-minute time window. Let's denote the counts for each protocol as follows: C_{TCP}: count of TCP packets, C_{UDP}: count of UDP packets, C_{ICMP}:

Table I: Summary of Datasets

Dataset	Year	Size	Duration	Format	Publicly Available	Studied	Contains Attack Traffic	Contains Legitimate Traffic	Attack Types
Booters [23]	2013	250GB	2 days	Packet	Yes	No	Yes	No	9 types of DDoS attacks
CIC DoS [24]	2017	5GB	24 hours	Packet	Yes	No	Yes	Yes	8 application layer DoS attack traces
DARPA [25]	1999	10GB	5 weeks	Packet	Yes	No	Yes	Yes	DoS, privilege escalation, probing attacks
KDD CUP 99 [26]	1998	<1GB	N/A	Connections	Yes	No	Yes	Yes	DoS, privilege escalation, probing attacks
NSL-KDD [27]	2009	<1GB	5 weeks	Not Specified	Yes	No	Yes	Yes	Same as KDD CUP 99
PU-IDS [23]	1998	N/A	N/A	N/A	Yes	No	Yes	Yes	Same as NSL-KDD
NDSec-1 [28]	2016	2GB	Not Specified	Packet	Yes	No	Yes	No	Brute force, DDoS, exploits, probe, spoofing
Kyoto 2006+ [29]	2006-2015	>100GB	>9 years	Bro IDS sessions	No	No	Yes	Yes	Various attacks against honeypots
Santa [30]	2014	N/A	N/A	N/A	No	No	Yes	Yes	(D)DoS, DNS amplification, heartbleed, port scans
SSENET-2011 [31]	2011	N/A	4 hours	N/A	No	No	Yes	Yes	DoS, port scans, etc
SENET-2014 [32]	2014	N/A	4 hours	N/A	No	No	Yes	Yes	botnet, flooding, port scan
TUIDS [33]	2012	N/A	21 days	Packet, Flow	No	No	Yes	Yes	botnet, DDoS
DDoS 2016 [34]	2016	<1GB	N/A	Textual packet summary	Yes	No	Yes	Yes	4 types of DDoS attacks
CICIDS 2017 [35]	2017	50GB	5 working days	Packet	Yes	Yes	Yes	Yes	DoS, DDoS, etc
ISCX 2012 [36]	2012	85GB	7 days	Packet	Yes	Yes	Yes	Yes	HTTP DoS, DDoS, network infiltration, SSH brute force
UNSW-NB15 [37]	2015	150GB	31 hours	Packet	Yes	Yes	Yes	Yes	DoS, exploits, fuzzers, etc
CSE-CIC-IDS2018 on AWS [35]	2018	500GB	2 days	Packet	Yes	Yes	Yes	Yes	Brute force, DoS, Botnet, DDoS attacks
DDoS 2019 [38]	2019	150GB	Not Specified	Packet	Yes	Yes	Yes	Yes	SYN Flood, UDP Flood, HTTP Flood, DNS Flood, etc
URG'16 [19]	2016	14GB	4 months	Netflow	Yes	Yes	Yes	Yes	DoS, Informations Gathering Attack, Probe and Botnet
LITNET-2020 [18]	2020	>1GB	10 months	Netflow	Yes	Yes	Yes	Yes	TCP SYN-flood, UDP-flood, HTTP-flood, etc

count of ICMP packets, C_{OTHER}: count of packets with other (IP) protocols. Next, we calculate the total number of packets in the one-minute time window, denoted as Tc:

$$Tc = C_{TCP} + C_{UDP} + C_{ICMP} + C_{OTHER} \quad (1)$$

Now we can compute the different protocol ratios R_p, where $p \in \{TCP, UDP, ICMP, OTHER\}$:

$$R_p = \frac{C_p}{Tc} \quad (2)$$

The "Protocol distribution" metric can be represented as a tuple or vector of the protocol ratios:

$$ProtocolDistribution = (R_{TCP}, R_{UDP}, R_{ICMP}, R_{OTHER}) \quad (3)$$

Where $R_{TCP} + R_{UDP} + R_{ICMP} + R_{OTHER} = 1$. These ratios represent the relative frequency of each protocol type within the one-minute time window, and their sum equals 1 as they together account for 100% of the packets in that window.

- *Entropy of source and destination ports*: This metric helps describing the diversity of ports used in the traffic and can potentially indicate the presence of attacks targeting specific ports. For each one-minute time window, we calculate the entropy of the port distribution using the formula:

$$H(P) = -\sum_i p_i \log p_i \quad (4)$$

Where $H(P)$ is the entropy of the port distribution, P is the set of unique ports and p_i is the probability that a packet uses port i.

We then plotted a graphical representation of these characteristics to see their evolution over time. We also compared these characteristics between the different datasets and with a CSP production traffic.

VI. Literature Dataset Quantification

In this section, we demonstrate the relevance of the previously introduced metrics on DDoS attack datasets from scientific literature. The focus is not to detail a specific detection technique, but rather to showcase the data that experts rely on to identify crisis situations. To achieve this, we first present how these metrics manifest on benign traffic, before showing how they can be used to characterize volumetric DDoS attacks. We then discuss the implications of these results at the scale of a CSP.

A. Normal flows characterization

Figure 1a presents the four aforementioned metrics for a "normal" network traffic — *i.e.* without DDoS attacks — on Monday's data from the CICIDS2017 dataset. This representation effectively exhibits the progression of the metrics under standard operational conditions. Upon examining the *Number of source IP addresses per destination IP address* metric, we observe a random distribution centered around an approximate average of 100 source IP per destination IP. The figure also unveils an initial "startup" phase at the beginning of the day, where 12 a.m. (GMT) corresponds to 9 a.m. in the timezone of the dataset.

Moreover, a "seasonal effect" or "cyclostationarity" [19] emerges in the early morning on the *Total number of packets observed* metric, surpassing 10,000 packets per second during the initial hour of connectivity before settling to an average of 1,000 packets per second subsequently. This seasonal effect is further manifested at the onset of the dataset for the *Protocol distribution* metric and the *entropy* calculation of the ports. In each instance, the metric under consideration adheres to a random distribution encircling an average value, punctuated by sporadic seasonal fluctuations.

The June 11th data from the ISCX IDS 2012 dataset is another example of a dataset without any trace of DDoS attack. An expert in the field will observe seasonal (daily) effects related to working hours in these records as shown in Figure 1b. From 9 a.m. to 6 p.m., the traffic has a distinctly different pattern compared to nighttime traffic, for each of the indicators. The maximum number of source IPs per destination is higher during the day than at night, and the protocol ratios differ. The total number of packets transmitted per minute and its variance over time are both noticeably different. Specific "events" can be observed at 6:00 a.m. and 6:00 p.m. regarding port entropy, as well as a higher average entropy during the day compared to nighttime. Once again, understanding the seasonal characteristics of network traffic appears to be important.

B. DDoS identification

1) CICIDS2017: Figure 2a focuses on the Wednesday data from the CIC IDS 2017 dataset, which contains four DoS attacks.

The metrics identified by the literature can highlight the attacks without substantial difficulty. Both Slowloris attacks are clearly visible in the *Total packets count* metric, which exhibits spikes an order of magnitude greater than the average value. The *Protocol distribution* metric conveys the same information. The potential challenge in identifying the attack based on these two metrics lies in the fact that they are scarcely distinguishable from the morning's seasonal peak. The *entropy* measurement of source ports, which is markedly distinct from the entropy measurement of destination ports during the attacks, serves as another notable marker. The familiarity with these three markers and the network's seasonal behaviors enables an expert to swiftly identify volumetric DDoS attack scenarios. DoS attacks 3 and 4 are not visible on the graphs, as they are not volumetric DDoS attacks targeting the network. Instead, they involve the use of low-volume HTTP traffic designed to put a heavy load on the server. This underscores the fact that CSPs cannot manage this type of problem at the global level of their network infrastructure. Consequently, each customer, depending on their specific use cases, must conduct a tailored analysis of their servers within the overall infrastructure in order to address these attacks in their cloud infrastructure.

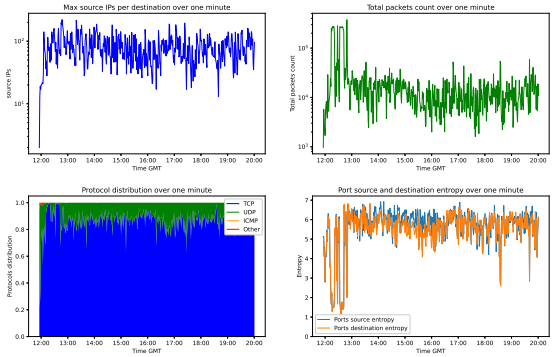
The CSE-CIC-IDS2018 dataset contains several DDoS attacks. On Tuesday, two DDoS attacks are present, DDoS attacks-LOIC-HTTP and DDoS-LOIC-UDP.

The second attack is clearly identified by the previously described metrics as shown by Figure 2b. It is distinctly visible in the *Total packets count per minute*, marked by a significant spike, which corresponds to a sudden variation in the protocol ratios (favoring UDP, given the nature of the attack) as seen in the *Protocol distribution* metric, and a fluctuation in the destination port entropy.

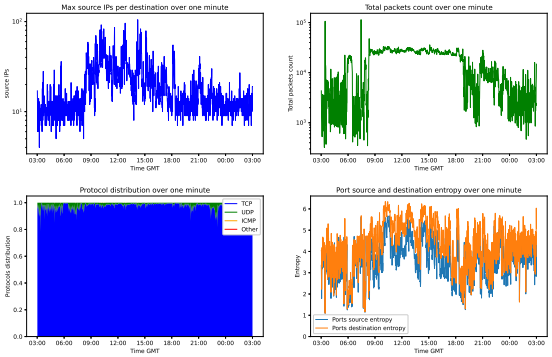
The first attack, on the other hand, does not involve any particular network traffic as it is a DDoS attack on a web server, specifically targeting the server's resources rather than the infrastructure as a whole.

2) DDoS 2019: Figure 3a presents the metrics measured for the CIC DDoS 2019 dataset on the second day. As a reminder, there are 12 DDoS attacks on that day, most of them being volumetric ones.

The three key metrics from the literature – *total packets per minute*, *Protocol distribution*, and *entropy* of incoming and outgoing ports – clearly indicate each of these volumetric attacks, demonstrating the relevance of these measures for DDoS detection. The only metric that is not relevant based on the

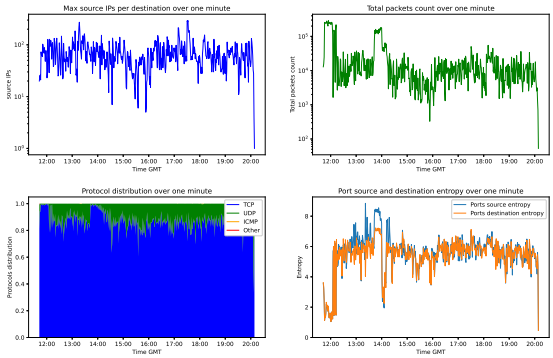


(a) CIC IDS 2017 Monday working hours.

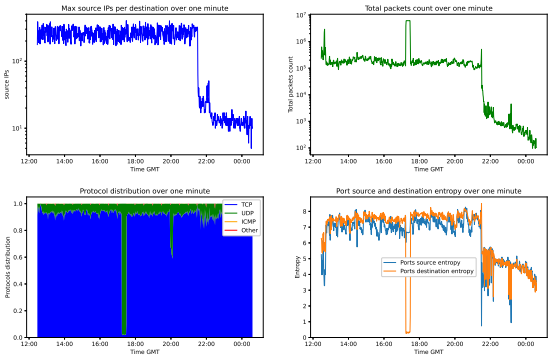


(b) ISCX IDS 2012 11 June.

Figure 1: Datasets with no DDoS attacks.



(a) CIC IDS 2017 Wednesday working hours.



(b) CSE CIC IDS 2018 Tuesday 20-02-2018.

Figure 2: Datasets that includes DDoS attacks.

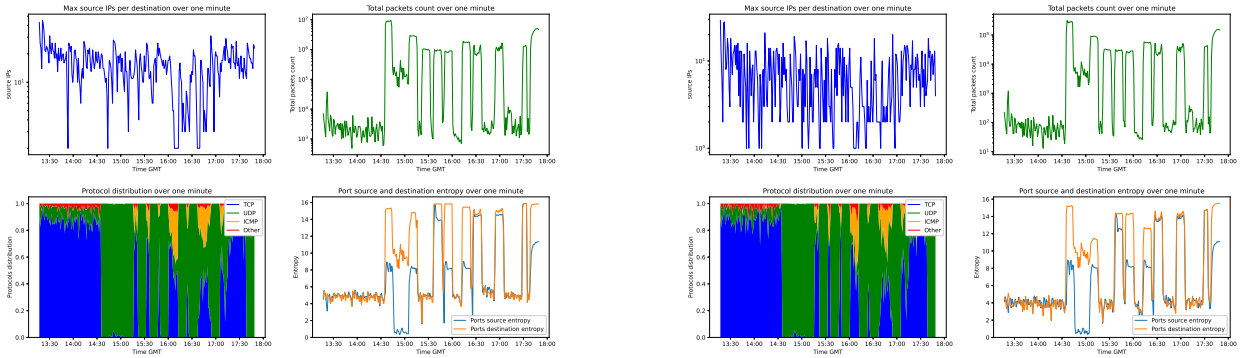
reported information is the *Max IP sources per destination* metric. This can be explained by the fact that during these attacks, no botnets were used, and IP spoofing was not implemented.

VII. Cloud Service Provider Quantification

To quantify a CSP traffic, we conducted on a CSP network a data capture spanning twelve hours, from 8 a.m. (GMT) to 20 p.m. (GMT), using a consistent methodology to track the temporal evolution of the selected metrics. The traffic was captured using packet sampling. This capture method is mandatory because of the scale of the observed traffic. Indeed, the observed infrastructure is composed of hundreds of thousands of servers hosting a large customer base of over 1.5 million users. The resulting network traffic reaches several terabytes per second. As a result, sampling techniques such as NetFlow [40] or sFlow [41] documented in the scientific literature are employed. These sampling methods enable the collection of a representative subset of the overall traffic, providing insights into traffic characteristics without the need for real-time monitoring of the entire flow. Furthermore, in line with the observations of Androulidakis *et al.* [42], the effectiveness of sampling relies solely on the sampling rate and not on the specific methodology employed. With an approximate sampling rate of 1/2000 for incoming traffic and 1/4000 for outgoing traffic, and considering the capability to observe billions of packets per second on the infrastructure, we have reasonable confidence in the statistical representativeness of the data, comparable to that extracted from the datasets mentioned in the literature. We focused exclusively on IPv4 flows, as the datasets predominantly consist of IPv4 traffic. Figure 4 presents the evolution of the obtained metrics throughout the capture period. It is noteworthy that the values remain largely homogeneous, despite the occurrence of around ten thousands DDoS attacks detected by our existing volumetric DDoS attack detection system. Unlike what has been observed in the literature datasets, these attacks have no discernible impact on the metrics. However, a slight upward trend is observed as the day progresses, which can be attributed to the previously mentioned seasonal effect.

VIII. Discussion

We have been able to demonstrate the evolution of the selected metrics in the literature datasets where they were not previously calculated. However, for the URG'16 [19] and LITNET-2020 [18] datasets, the authors had already performed this analysis in their respective articles. We encourage the reader to refer to Figures 1 and 4 for URG'16, and Figures 5, 6, and 7 for LITNET-2020, to gain further insights into these specific datasets. The various illustrations of the primary metrics, derived from the existing body of research, demonstrate that the essential information for detecting



(a) No sampling.

(b) 1/32 packet sampling rate.

Figure 3: CIC DDoS 2019 on 1st december with and without sampling.

volumetric DDoS attacks becomes discernible once the metrics are computed. Upon initial analysis, it may appear reasonable to anticipate comparable findings when observing these metrics on CSP backbones. Nonetheless, a prominent difference exists: the volume of packets in transit is significantly greater than that of the datasets. For instance, the network traffic observed on the backbone of the considered CSP amounts to approximately 2,5 billions packets per second, whereas the datasets' volume never surpasses 40,000 packets per second.

This change in scale, in terms of packet volume in transit, inevitably impacts the computation of metrics defined by the literature and presented in section V and illustrated in section VI. First, as stated in section VII, it is not feasible to capture all traffic. Therefore, the statistics are based on a sample of packets. The use of samples for statistical production is a well-documented field, even in the context of network metrics. However, we believe it is necessary to verify whether this sampling is not the root of the problem. Second, this change in scale affects the ratio of healthy data volume to the volume of data generated for conducting a volumetric attack. It is critical to ascertain whether this factor is not the source of a dilution of the sought signal, which becomes weak in the face of regular noise.

To examine the first hypothesis, we employed a sampling methodology on the state-of-the-art datasets; subsequently, we computed the identical metrics as those on the original datasets, but using this sampled version. Due to space constraints, we include here merely one of the figures obtained: the CIC DDoS 2019 01-12 dataset, using a sampling rate of 1/32.

Figure 3b illustrates that properly executed sampling (randomly selecting one packet out of n to sample) does not lead to a loss of information within the aggregated data. This Figure represents the same datasets as the one depicted in Figure 3a, with a sampling rate of 1/32. The insights underscored in Figure 3a are unequivocally identifiable in Figure 3b.

We have successfully reproduced these results on a variety of other datasets, which we cannot extensively discuss in this article due to space constraints. However, these additional datasets consistently produce the same findings. Furthermore, the study by Boin *et al.* [43] has established that many seasonal effects are observable at the scale of a CSP. It appears, therefore, that there are no impediments to the well-known techniques in scientific literature being applicable in the context of network analysis.

Nevertheless, the sheer volume of network traffic raises another issue: that of weak signals. Indeed, even during a volumetric DDoS attack, a CSP's customer may generate network traffic that is negligible compared to the overall network traffic. Identifying this weak signal against legitimate network usage peaks could be the genuine obstacle hindering the implementation of DDoS detection solutions at the scale of a CSP.

To corroborate this hypothesis of weak signals, we conducted another experiment. By blending 100 times the benign traffic from Figure 1a with one instance of DDoS traffic from Figure 3a, we generated the metrics depicted in Figure 5. In this scenario, it becomes apparent to the observer that the sought-after information is diluted within the total volume. The volumetric attack has indeed become a weak signal, submerged within an overall benign volume.

Furthermore, in the studies conducted by Damasevicius *et al.* [18] and Maciá-Fernández *et al.* [19], it can be observed that although the impact of DDoS attacks on these metrics remains visible, it diminishes as the size of the infrastructure increases.

This analysis offers a crucial understanding of the challenges faced in detecting volumetric DDoS attacks within the setting of a CSP. The primary issue stems from the weak signal that these attacks present within the enormous volume of network traffic, which proves to be a significant obstacle. It becomes clear that while traditional metrics are successful in smaller, literature-derived datasets, they may struggle to effectively identify these attacks in the vast expanse of everyday network operations at such a large scale. Therefore, future research should focus on developing innovative methods and metrics that can address these unique challenges inherent to the CSP environment.

IX. CONCLUSION

In this paper, we have highlighted the differences we observe between the volumetric DDoS datasets proposed in

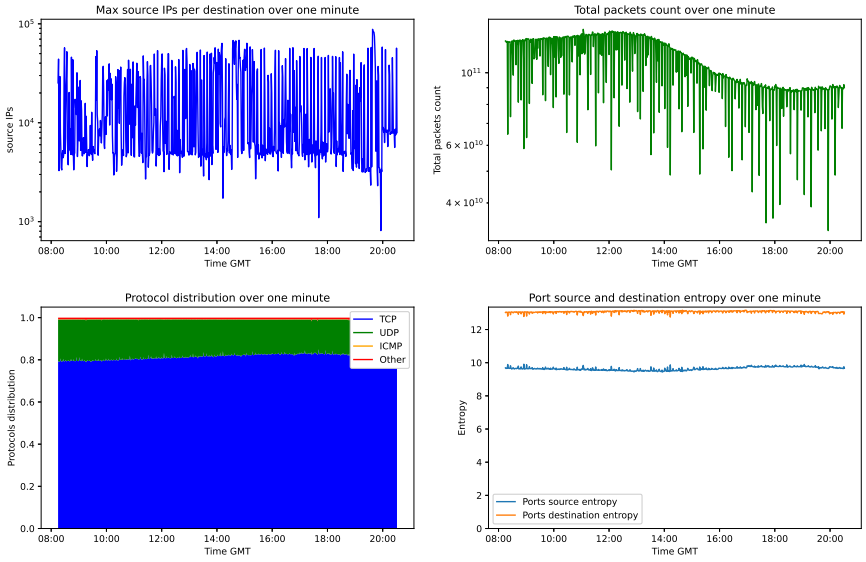


Figure 4: Metrics observed on the captured traffic of a CSP backbone.

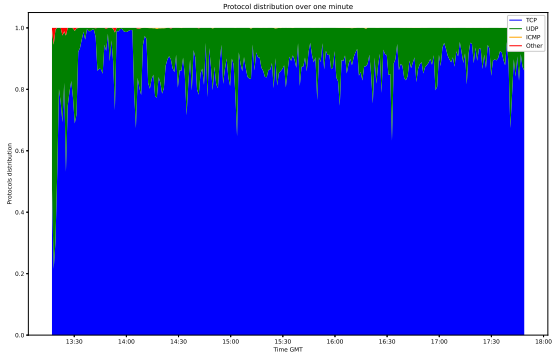


Figure 5: Protocol distribution observed when blending CIC DDoS 2019 attacks and CIC DDoS 2017 benign traffic with a 1/100 ratio.

the scientific literature and what we observe on the network backbone of the CSP we work for.

We have explained that the volume of data in transit on the network backbones of CSPs necessitates the implementation of sampling techniques.

We have also empirically established that these sampling techniques do not have detrimental effects on the establishment of volumetric DDoS detection metrics.

However, we have observed that the volumes of data in transit on these backbones are orders of magnitude higher than those in the datasets used in the literature.

Finally, we have shown that the gargantuan volumes of data lead to a difficulty not identified by the literature:

Volumetric attacks directed at one of the CSP's clients have a very limited impact on the metrics traditionally used to detect them, when collected from the traffic managed by the CSP.

Identifying these attacks by analyzing the global traffic of the CSP is akin to searching for a weak signal, particularly in the face of the seasonal effects of traffic.

This situation is therefore not comparable to that observed in the datasets proposed and used by the literature.

In accordance with these observations, we propose to direct our future work to:

- Provide a means of testing DDoS detection techniques proposed by academic research on datasets that reflect the metrics we observe on the network backbones of CSPs. For obvious reasons of confidentiality, as well as data volume, it is not possible to present industry-derived data, which is why such datasets are not available in the literature. However, given legal and technical constraints, we believe it is reasonable to consider the use of traffic generators.
- Demonstrate how the signals exploited by volumetric DDoS detection techniques can be adapted to report volumetric attacks, even when they constitute only a weak signal of the network traffic being analyzed. Or, if this proves impossible, to propose new metrics.

References

[1] G. Somani, M. S. Gaur, D. Sanghi, M. Conti, and R. Buyya, "Ddos attacks in cloud computing: Issues, taxonomy, and future directions," *Computer Communications*, vol. 107, pp. 30–48, 2017.

[2] S. Greenstein, "The aftermath of the dyn ddos attack," *IEEE Micro*, vol. 39, no. 4, pp. 66–68, 2019.

[3] H. A. Herrera, W. R. Rivas, and S. Kumar, "Evaluation of internet connectivity under distributed denial of service attacks from botnets of varying magnitudes," in *2018 1st International Conference on Data Intelligence and Security (ICDIS)*. IEEE, 2018, pp. 123–126.

[4] L. Qian, Z. Luo, Y. Du, and L. Guo, "Cloud computing: An overview," in *Cloud Computing: First International Conference, CloudCom 2009, Beijing, China, December 1-4, 2009. Proceedings 1*. Springer, 2009, pp. 626–631.

[5] M. McNutt, "Reproducibility," pp. 229–229, 2014.

[6] S. Behal and K. Kumar, "Trends in validation of ddos research," *Procedia Computer Science*, vol. 85, pp. 7–15, 2016.

[7] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, "A survey of network-based intrusion detection data sets," *Computers & Security*, vol. 86, pp. 147–167, 2019.

[8] M. Małowidzki, P. Berezinski, and M. Mazur, "Network intrusion detection: Half a kingdom for a good dataset," in *Proceedings of NATO STO SAS-139 Workshop, Portugal*, 2015.

[9] B. B. Gupta and O. P. Badve, "Taxonomy of dos and ddos attacks and desirable defense mechanism in a cloud computing environment," *Neural Computing and Applications*, vol. 28, pp. 3655–3682, 2017.

[10] A. Toh, "Azure ddos protection—2021 q3 and q4 ddos attack trends," https://azure.microsoft.com/en-us/blog/azure-ddos-protection-2021-q3-and-q4-ddos-attack-trends/, 2022.

[11] Y. Khalidi, "How microsoft builds its fast and reliable global network," https://azure.microsoft.com/en-au/blog/how-microsoft-builds-its-fast-and-reliable-global-network/, 2017.

[12] M. Darwish, A. Ouda, and L. F. Capretz, "Cloud-based ddos attacks and defenses," in *International Conference on Information Society (i-Society 2013)*. IEEE, 2013, pp. 67–71.

[13] C. Thomas, V. Sharma, and N. Balakrishnan, "Usefulness of darpa dataset for intrusion detection system evaluation," in *Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security 2008*, vol. 6973. SPIE, 2008, pp. 164–171.

[14] L. Dhanabal and S. Shantharajah, "A study on nsl-kdd dataset for intrusion detection system based on classification algorithms," *International journal of advanced research in computer and communication engineering*, vol. 4, no. 6, pp. 446–452, 2015.

[15] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," *Ieee communications surveys & tutorials*, vol. 16, no. 1, pp. 303–336, 2013.

[16] A. Gharib, I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "An evaluation framework for intrusion detection dataset," in *2016 International Conference on Information Science and Security (ICISS)*. IEEE, 2016, pp. 1–6.

[17] C. G. Cordero, E. Vasilomanolakis, A. Wainakh, M. Mühlhäuser, and S. Nadjm-Tehrani, "On generating network traffic datasets with synthetic attacks for intrusion detection," *ACM Transactions on Privacy and Security (TOPS)*, vol. 24, no. 2, pp. 1–39, 2021.

[18] R. Damasevicius, A. Venckauskas, S. Grigaliunas, J. Toldinas, N. Morkevicius, T. Aleliunas, and P. Smuikys, "Litnet-2020: An annotated real-world network flow dataset for network intrusion detection," *Electronics*, vol. 9, no. 5, p. 800, 2020.

[19] G. Maciá-Fernández, J. Camacho, R. Magán-Carrión, P. García-Teodoro, and R. Therón, "Ugr '16: A new dataset for the evaluation of cyclostationarity-based network idss," *Computers & Security*, vol. 73, pp. 411–424, 2018.

[20] C. O. Camargo, E. R. Faria, B. B. Zarpelão, and R. S. Miani, "Qualitative evaluation of denial of service datasets," in *Proceedings of the XIV Brazilian Symposium on Information Systems*, 2018, pp. 1–8.

[21] "1999 DARPA Intrusion Detection Evaluation Dataset | MIT Lincoln Laboratory — ll.mit.edu," https://www.ll.mit.edu/r-d/datasets/1999-darpa-intrusion-detection-evaluation-dataset, [Accessed 16-May-2023].

[22] "The CAIDA "DDoS Attack 2007" Dataset — caida.org," https://www.caida.org/catalog/datasets/ddos-20070804_dataset/, [Accessed 16-May-2023].

[23] R. Singh, H. Kumar, and R. Singla, "A reference dataset for network traffic activity based intrusion detection system," *International Journal of Computers Communications & Control*, vol. 10, no. 3, pp. 390–402, 2015.

[24] H. H. Jazi, H. Gonzalez, N. Stakhanova, and A. A. Ghorbani, "Detecting http-based application layer dos attacks on web servers in the presence of sampling," *Computer Networks*, vol. 121, pp. 25–36, 2017.

[25] R. P. Lippmann, D. J. Fried, I. Graf, J. W. Haines, K. R. Kendall, D. McClung, D. Weber, S. E. Webster, D. Wyschogrod, R. K. Cunningham *et al.*, "Evaluating intrusion detection systems: The 1998 darpa off-line intrusion detection evaluation," in *Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00*, vol. 2. IEEE, 2000, pp. 12–26.

[26] [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[27] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," in *2009 IEEE symposium on computational intelligence for security and defense applications*. Ieee, 2009, pp. 1–6.

[28] F. Beer, T. Hofer, D. Karimi, and U. Bühler, "A new attack composition for network security," in *10. DFN-Forum Kommunikationstechnologien*. Gesellschaft für Informatik eV, 2017.

[29] J. Song, H. Takakura, Y. Okabe, M. Eto, D. Inoue, and K. Nakao, "Statistical analysis of honeypot data and building of kyoto 2006+ dataset for nids evaluation," in *Proceedings of the first workshop on building analysis datasets and gathering experience returns for security*, 2011, pp. 29–36.

[30] C. Wheelus, T. M. Khoshgoftaar, R. Zuech, and M. M. Najafabadi, "A session based approach for aggregating network traffic data–the santa dataset," in *2014 IEEE International Conference on Bioinformatics and Bioengineering*. IEEE, 2014, pp. 369–378.

[31] A. Vasudevan, E. Harshini, and S. Selvakumar, "Ssenet-2011: a network intrusion detection system dataset and its comparison with kdd cup 99 dataset," in *2011 second asian himalayas international conference on internet (AH-ICI)*. IEEE, 2011, pp. 1–5.

[32] S. Bhattacharya and S. Selvakumar, "Ssenet-2014 dataset: A dataset for detection of multiconnection attacks," in *2014 3rd International Conference on Eco-friendly Computing and Communication Systems*. IEEE, 2014, pp. 121–126.

[33] P. Gogoi, M. H. Bhuyan, D. Bhattacharyya, and J. K. Kalita, "Packet and flow based network intrusion dataset," in *Contemporary Computing: 5th International Conference, IC3 2012, Noida, India, August 6-8, 2012. Proceedings 5*. Springer, 2012, pp. 322–334.

[34] M. Alkasassbeh, G. Al-Naymat, A. B. Hassanat, and M. Almseidin, "Detecting distributed denial of service attacks using data mining techniques," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 1, 2016.

[35] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization." *ICISSp*, vol. 1, pp. 108–116, 2018.

[36] A. Shiravi, H. Shiravi, M. Tavallaee, and A. A. Ghorbani, "Toward developing a systematic approach to generate benchmark datasets for intrusion detection," *computers & security*, vol. 31, no. 3, pp. 357–374, 2012.

[37] N. Moustafa and J. Slay, "Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set)," in *2015 military communications and information systems conference (MilCIS)*. IEEE, 2015, pp. 1–6.

[38] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (ddos) attack dataset and taxonomy," in *2019 International Carnahan Conference on Security Technology (ICCST)*. IEEE, 2019, pp. 1–8.

[39] S. Sharma, S. K. Sahu, and S. K. Jena, "On selection of attributes for entropy based detection of ddos," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2015, pp. 1096–1100.

[40] B. Claise, "Cisco systems netflow services export version 9," Tech. Rep., 2004.

[41] P. Phaal, S. Panchen, and N. McKee, "Inmon corporation's sflow: A method for monitoring traffic in switched and routed networks," Tech. Rep., 2001.

[42] G. Androulidakis, V. Chatzigiannakis, S. Papavassiliou, M. Grammatikou, and V. Maglaris, "Understanding and evaluating the impact of sampling on anomaly detection techniques," in *MILCOM 2006-2006 IEEE Military Communications conference*. IEEE, 2006, pp. 1–7.

[43] C. Boin, X. Guillaume, G. Grimaud, T. Groléat, and M. Hauspie, "One Year of DDoS Attacks Against a Cloud Provider: an Overview," in *4th International Conference on Advances in Computer Technology, Information Science and Communications*, Suzhou, China, Apr. 2022. [Online]. Available: https://hal.science/hal-03655003