Can Blocklists Explain Darknet Traffic?

Damiano Ravalico[†], Rodolfo Valentim[‡], Martino Trevisan[†], Idilio Drago[‡]
[†]University of Trieste, [‡]University of Turin
first.last@{phd.units,dia.units,unito}.it

***Abstract*—Darknets are IP addresses that function as passive probes, recording all received packets without hosting services. The traffic they capture, being unsolicited, makes darknets akin to "network telescopes". Traces collected on darknets aggregate multiple events useful for cybersecurity, like network scans and exploit attempts. Yet, the mix of heterogeneous events observed from darknets poses significant challenges to those who must understand darknet traffic.**

Here we face the question of whether new darknet deployments provide novel and useful information when compared to *public blocklists*. Multiple Cyber Threat Intelligence (CTI) sources publish lists of IP addresses that perform malicious activities, from simple automated scans to SPAM and phishing campaigns. They represent a valuable resource for network administrators, helping to block cyberattacks. Built with a combination of multiple sensors — including darknets and honeypots — these lists could explain the traffic seen on other darknets, thus simplifying the search for relevant events in independent darknet deployments.

We thus investigate to what extent open blocklists explain darknet traffic. By crawling hundreds of CTI sources providing blocklists, we first notice how these lists are often incomplete or slowly updated. Traffic seen in our darknet deployment is hardly explained by the blocklists, even when considering only the most prominent scan attempts, and ignoring events such as backscattering. Our preliminary results suggest that blocklists can be of great use for seeding the explanation of darknet traffic, by giving context for the activity of a few IP addresses. Yet, more addresses with similar behaviour are observed in the darknet and could be used to enrich and complement the blocklists.

***Index Terms*—Darknets; Blocklists; cybersecurity**

I. INTRODUCTION

Darknets (or network telescopes) consist of IP addresses publicly announced on the Internet routing, yet hosting no services [1], [2]. Darknets receive a mixture of traffic composed of network scans, backscattering, and the outcomes of misconfigured hosts. The intrinsic unsolicited nature of such traffic makes Darknets useful for providing insights into cybersecurity events. Yet, analyzing darknet traffic proves to be a daunting task. The diverse range of events and the huge numbers of sources observed within darknets complicate efforts to put the traffic into useful perspectives. Moreover, whereas it has been shown that different darknets observe complementary events [3], the deployments of new darknets require the allocation of (perhaps scarce) IP addresses to the tasks.

These challenges raise questions about whether new darknet deployments observe novel events that are worth the costs of their deployments. In particular, Cyber Threat Intelligence (CTI) companies and public entities release multiple open lists of IP addresses performing malicious activities. These lists are populated using a variety of complementary approaches. Examples include lists composed of IP addresses sending out SPAM emails, IP addresses hosting and distributing malware software, as well as IP addresses observed in large darknet and honeypot deployments run by security companies. These lists are then consolidated and distributed as blocklists [4], which network administrators employ to block malicious traffic, thus preventing or slowing down cyberattacks.

978-3-903176-64-5 ©2024 IFIP

We here face the question of whether public blocklists can be used to explain darknet traffic. A positive answer to this question would mean that novel darknet deployments are superfluous, providing information already available on CTI sources. To answer this question, we rely on data collected from a darknet deployed in an academic network for a full month. We then crawl hundreds of public blocklists released in the same period. We first evaluate the IP addresses observed in the public lists. Here we confirm and reappraise findings of previous works [4], [5]: These lists are very heterogeneous, and their update policies and frequency vary greatly across providers.

We then select a subset of 28 lists and compare the reported IP addresses with the traffic observed in our darknet. To filter out sporadic events, we focus only on the most active IP addresses observed in the darknet. Our initial results show some intersection between the darknet and IP addresses reported in blocklists as expected. However, the majority of IP addresses seen in the darknet are never reported in the blocklists. When considering only IP addresses sending at least 5 packets to the darknet, around 3 % of the IP addresses observed in the darknet are reported. The overlap between darknet and blocklists is proportional to the list sizes, but a single blocklist can only explain an insignificant portion of the darknet traffic.

These are the initial results of our effort to automate the analysis of darknet traffic. They show that new darknet deployments do have the potential to provide novel insights into malicious traffic, which are not yet visible on public blocklists. For that, we believe the blocklists provide valuable seeds to start the analysis, giving hints that explain the traffic of IP addresses contacting the darknet. Our plan for future work is to combine blocklists with approaches to cluster darknet traffic, such as [6], [1], thus building methods to augment and complement blocklists automatically.

TABLE I: Darknet dataset overview.

Volume	9.5 GB
Packets	117 778 528
TCP	94.4 %
UDP	5.5 %
IP Addresses	670 754
IP Addresses (Filtered)	257 855

II. Dataset

A. Darknet

We rely on data from a darknet composed of IPv4 addresses allocated to a university in Italy. It is formed by two /24 networks, with non-continuous addresses. The IPv4 ranges are kept private following requests of the research institutions running the networks. We set up a network probe to capture the traffic arriving at the allocated addresses, recording the full packets. The probe obfuscates IPv4 prefixes of the darknets (i.e., destination IP addresses). We perform analyses using data collected during 1 month, from the 1st of March to the 1st of April 2024. In Table I, we report the most salient features of the dataset.

We filter packets to focus on the most salient events seen in the darknet. First, we discard TCP packets which are not *pure* SYNs packets. Indeed, packets with the ACK, RST or FIN flag set are typically backscattering traffic — i.e., traffic sent by a victim in response to packets received with a spoofed source IP address (spoofed source belonging to the darknet range). Second, as done in previous work [6], [1], we filter out IP addresses sending less than 5 packets. The resulting dataset includes 257 855 IP address, which we seek in the public blocklists.

B. Blocklists

We use various publicly available blocklists sourced from the `FilterLists` online aggregator [7]. This platform hosts over 300 public lists spanning diverse fields, encompassing malware or phishing websites, popular ad- and tracker-blocker lists, and notably, several lists containing sets of malicious IP addresses associated with suspicious traffic, scans, or attacks on specific web services. `FilterLists` categorizes these lists based on their syntax (such as IP set or URL list) and the type of content they specialize in.

For our analyses, we focus on lists containing IP addresses or IP ranges (IPv4 only) associated with malware distribution. These lists are typically formatted as "host files", where each row represents an IP address. In this initial analysis, we have selected 28 blocklists from various maintainers. The full list of these blocklists is provided in the Appendix.

III. Initial Results

We now present our initial findings. First, we quantify how many darknet addresses are found in the lists, and, then, we measure the visibility a darknet offers on the addresses contained in specific lists. Finally, we provide a more detailed analysis of the intersection between the darknet and the blocklists, in particular evaluating the type of traffic seen in the darknet for IP addresses of each blocklist.

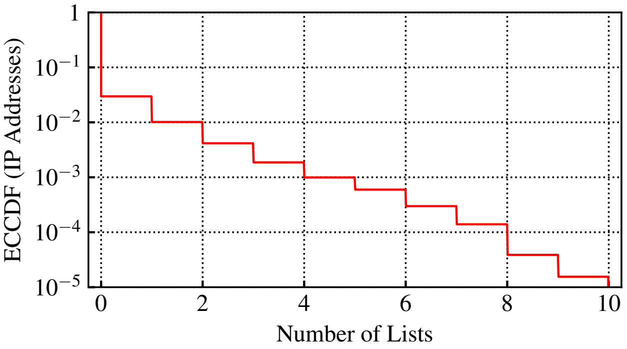

Fig. 1: Number of lists in which IP addresses are found.

A. How many darknet addresses do appear in a list?

In Figure 1, we plot the Empirical Cumulative Distribution Function (ECDF) of the number of lists each IP address observed in the darknet appears in. We use the log scale on the y-axis to emphasize the tail of the distribution — i.e., addresses appearing in many lists. Indeed, 36 addresses appear in 8 or more lists. The vast majority of IP addresses are not seen in any list. Specifically, among the total of 257 855 filtered IP addresses, 250 199 addresses are not found in any list. This result indicates that the perspective given by these lists does not provide a comprehensive representation of traffic captured in the darknets.

Only 2.96 % of the unique addresses seen in the darknet can be found in the blocklists. Of these, 5040 can be found in one list, 1545 in two lists and 1071 in more than two lists. Our analysis also confirms that the overlap among some blocklists is not negligible, as pointed out by previous studies [4]. Indeed, not shown for the sake of space, we observe that lists are very heterogeneous across providers.

The low percentage of IP addresses in the blocklists can be partly explained by the fact that darknets observe a large number of non-malicious scanners [6], [1]. Blocklists are not expected to report these IP addresses. Yet, the percentage of traffic explained by the blocklists remains small even when filtering out traffic from well-known scanners.

B. Darknet visibility

We now quantify to what extent the darknet allows observing the traffic originating from the set of known malicious actors included in different blocklists. Figure 2 reports on the x-axis the size of each list and, on the y-axis, the size of the overlap with addresses observed in the darknet. Both x- and y-axes use logarithmic scales, given the great differences in list size. The top three largest lists are respectively "BlockConvert (IPs)" (160 278 addresses), "gnX Threat Intelligence" (115 101

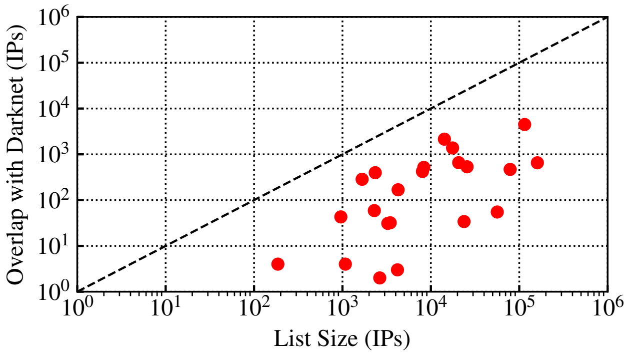

Fig. 2: List size and overlap with addresses seen in the darknet.

addresses) and “Inversion DNSBL” (78 811 addresses). Surprisingly, the absolute overlap with the darknet for the first and last of these lists is quite low, at 653 and 468 addresses, respectively. Conversely, the list “gnX Threat Intelligence” exhibits the highest overlap, with 4475 addresses seen in the darknet. Beyond “gnX Threat Intelligence”, the lists with the highest overlap are “CINS Army Bad Guys” (2141 addresses out of 14 231) and “Firehol Level 3” (1376 addresses out of 17 616). These lists that present the highest overlap are also the lists that generate the most traffic in terms of the number of packets. Furthermore, “BlockConvert (IPs)” and “Inversion DNSBL”, despite the very low overlap, fall into the lists that generate a lot of traffic. This suggests that lists report mostly top-talkers.

C. Most targeted destination ports

Darknet traffic can be used to study the traffic monitored by these lists. To offer an example in this direction, we investigate the most targeted destination TCP and UDP ports. Figure 3 breaks down the traffic of the five lists discussed in the previous section by the top six contacted ports (they are all TCP ports). Each column of the heatmap refers to a list and a cell reports the percentage of packets to the given port over the total traffic of the list. The last column reports the overall darknet traffic to the top six ports. All of them are well-known or registered, but (not shown in the figure) we notice that some lists generate a lot of traffic to ephemeral ports. Overall, the most contacted port is 23 (Telnet), but each list exhibits different behaviour. Interestingly, among the most contacted ports, we find 2000 (Cisco Skinny Client Control Protocol), which is especially targeted by hosts in the “gnX Threat Intelligence” list, containing nodes likely to be compromised/infected. Differently, IP addresses listed in the “Inversion DNSBL” list (servers hosting malware according to the Safe Browsing API) perform scans on port 22 (Secure Shell). Finally, addresses in the “Firehol Level 3” list (which tracks attacks, spyware, and viruses) target especially port 3389 (Windows Remote Desktop).

In sum, the darknet allows observing the different behaviour of addresses from each list. As the addresses observed in the darknet are different from those in the blocklist, we believe the blocklist can be used as seeds to tag darknet traffic.

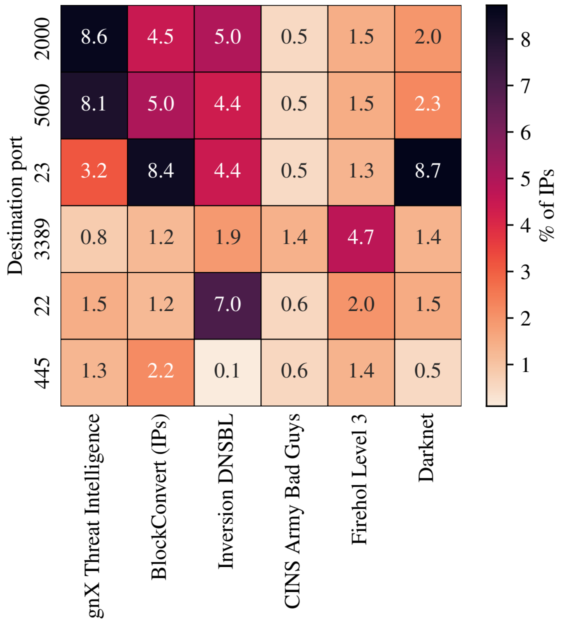

Fig. 3: Top-six TCP ports, separately for five lists and overall.

IV. Next Steps

Our initial results show that public lists can be successfully used to gain insights into a fraction of IP addresses contacting a darknet. Although we used a non-exhaustive set of blocklists, we find $\approx 3\,\%$ of the darknet addresses. Our current efforts include the collection of a larger set of lists and extending the temporal scope of the analyses.

Our ultimate goals are twofold. Firstly, we seek to investigate whether new and distributed darknets provide a means to increase the reliability of blocklists. If proven effective, we can use new darknets to systematically identify and remove obsolete entries from these lists. This process would help to refine blocklists and increase their applicability. Secondly, we aim to characterize addresses seen in darknets but which are not listed in any blocklist and automatically gain insights into their behaviour and intentions. In this scenario, blocklists serve as an important source of knowledge, guiding the identification of potentially malicious addresses that may have been overlooked currently.

Acknowledgments

The research leading to these results has been funded by the projects ACRE (AI-Based Causality and Reasoning for Deceptive Assets - 2022EP2L7H) and COMPACT (Compressed features and representations for network traffic analysis in centralized and edge Internet architectures — 2022M2Z728), funded by the European Union - Next Generation EU within the PRIN 2022 program (D.D. 104 - 02/02/2022 Ministero dell’Università e della Ricerca). This manuscript reflects only

the authors' views and opinions and the Ministry cannot be considered responsible for them.

References

[1] L. Gioacchini, L. Vassio, M. Mellia, I. Drago, Z. Houidi, and D. Rossi, "i-DarkVec: Incremental Embeddings for Darknet Traffic Analysis," *ACM Transactions on Internet Technology*, vol. 23, no. 3, pp. 1–28, 2023.

[2] M. Kallitsis, R. Prajapati, V. Honavar, D. Wu, and J. Yen, "Detecting and Interpreting Changes in Scanning Behavior in Large Network Telescopes," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 3611–3625, 2022.

[3] F. Soro, I. Drago, M. Trevisan, M. Mellia, J. Ceron, and J. J. Santanna, "Are darknets all the same? on darknet visibility for security monitoring," in *2019 IEEE international symposium on local and metropolitan area networks (LANMAN)*, pp. 1–6, IEEE, 2019.

[4] Á. Feal, P. Vallina, J. Gamba, S. Pastrana, A. Nappa, O. Hohlfeld, N. Vallina-Rodriguez, and J. Tapiador, "Blocklist babel: On the transparency and dynamics of open source blocklisting," *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 1334–1349, 2021.

[5] A. Sjösten, P. Snyder, A. Pastor, P. Papadopoulos, and B. Livshits, "Filter list generation for underserved regions," in *Proceedings of The Web Conference 2020*, WWW '20, (New York, NY, USA), p. 1682–1692, Association for Computing Machinery, 2020.

[6] L. Gioacchini, L. Vassio, M. Mellia, I. Drago, Z. B. Houidi, and D. Rossi, "Darkvec: Automatic analysis of darknet traffic with word embeddings," in *Proceedings of the 17th International Conference on emerging Networking EXperiments and Technologies*, pp. 76–89, 2021.

[7] C. M. Barrett, "Filterlists." https://filterlists.com/.

Appendix

For the sake of completeness, the following table reports the set of 28 public lists that we use in our study (downloaded using the FilterLists online API [7]).

- Binary Defense
- BlockConvert
- Blocklist.DE
- CINS Army Bad Guys
- Nordic Filters
- DangerRulezSK Brute Force Blocker
- EmergingThreats Block IPs
- EmergingThreats Compromised IPs
- Firehol Level 1
- Firehol Level 2
- Firehol Level 3
- Greensnow Blocklist
- IPsum Level 4
- IPsum Level 5
- IPsum Level 6
- ISX Solutions Blocklist
- Inversion DNSBL
- Maltrail - Parking sites
- Mirai Tracker
- MyIP Blacklist
- SecLists (Careto IPs by Kaspersky)
- gnX Threat Intelligence
- hpHosts EMD (IPs)
- hpHosts EXP (IPs)
- hpHosts FSA (IPs)
- pfBlockerNG - MS-1
- pfBlockerNG - MS-3
- Urlhaus-filter